Acquisition Directorate

Research & Development Center

Report No. CG-D-06-13

Intercomparison of U.S. Ballast Water Test Facilities

Final Report

November 2012

Homeland Security

N O T I C E

This document is disseminated under the sponsorship of the Department of Homeland Security in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

This report does not constitute a standard, specification, or regulation.

Mr. Timothy R. Girton
Technical Director
United States Coast Guard
Research & Development Center
1 Chelsea Street
New London, CT 06320

Technical Report Documentation Page

1. Report No. CG-D-06-13	**2. Government Accession Number**	**3. Recipient's Catalog No.**
4. Title and Subtitle Intercomparison of U.S. Ballast Water Test Facilities - Final Report		**5. Report Date** November 2012 **6. Performing Organization Code** Project No. 4101
7. Author(s) Lisa A. Drake, Timothy P. Wier, Jonathan F. Grant, Evan W.J. Parson, Edward J. Lemieux		**8. Performing Report No.** R&DC UDI #1022
9. Performing Organization Name and Address Naval Research Laboratory Chemistry Division 4555 Overlook Ave., SW Washington, DC 20375	U.S. Coast Guard Research and Development Center 1 Chelsea Street New London, CT 06320	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** Contract # HSCG32-09-X-R00023
12. Sponsoring Organization Name and Address U.S. Department of Homeland Security Commandant (CG-OES-3) United States Coast Guard 2100 Second St. SW Washington, DC 20593-0001		**13. Type of Report & Period Covered** Final **14. Sponsoring Agency Code** Commandant (CG-OES-3) U.S. Coast Guard Headquarters Washington, DC 20593-0001

15. Supplementary Notes

The R&D Center's technical point of contact is Ms. Gail Roderick , 860-271-2658, email: Gail.E.Roderick@uscg.mil

16. Abstract (MAXIMUM 200 WORDS)

The spread of aquatic nuisance species (ANS) via ships' ballast water has been studied for decades, and as the number of economically and environmentally destructive biological invasions rose, it became clear that national and international policies governing ballast water discharges were needed. The International Maritime Organization's International Convention for the Control and Management of Ships' Ballast Water and Sediments, adopted in 2004 but not yet entered into force, set a discharge standard to reduce the transport and delivery of potential ANS. Concurrently in the United States, the U.S. Coast Guard developed and finalized a discharge standard for living organisms in three size classes: ≥ 50 µm; ≥ 10 µm and < 50 µm; and < 10 µm. As vendors develop ballast water management systems (BWMSs) to meet the standards, they must be evaluated to determine their efficacy. The goals of this "Intercomparison Project" were to determine (1) to what degree two independent, domestic ballast water test facilities could comparably evaluate the performance of the same BWMS and (2) to identify and quantify the variability between the facilities. The goals were successfully met. The project, the BWMS testing, the resulting data, and recommendations to reduce variability among test facilities are discussed.

17. Key Words Aquatic nuisance species, ballast water treatment, biological invasions, invasive species, shipping	**18. Distribution Statement** Distribution Statement A: Approved for public release; distribution is unlimited.		
19. Security Class (This Report) UNCLAS//Public	**20. Security Class (This Page)** UNCLAS//Public	**21. No of Pages** 138	**22. Price**

(This page intentionally left blank.)

EXECUTIVE SUMMARY

The anthropogenic spread of aquatic nuisance species (ANS) via ships' ballast water has been studied and documented for decades, and as the number of economically and environmentally destructive biological invasions rose, it became increasingly clear that national and international policies governing ballast water discharges were needed. The International Maritime Organization's (IMO) International Convention for the Control and Management of Ships' Ballast Water and Sediments, adopted in 2004 but not yet entered into force, set a discharge standard to reduce the transport and delivery of potential ANS. Concurrently in the United States, the U.S. Coast Guard (USCG) developed and finalized a discharge standard for living organisms in three size classes: ≥ 50 µm; ≥ 10 µm and < 50 µm; and < 10 µm. In partnership with the Environmental Protection Agency (EPA), USCG signed a memorandum of agreement to develop a protocol for evaluating ballast water management systems (BWMSs) using land-based test facilities (TFs). The primary focus of that joint effort is to verify claims made by technology vendors regarding BWMS biological treatment performance, although other factors such as reliability, cost, and safety are also considered. With the input of stakeholders and technical panels, and after undergoing multiple EPA reviews and validation by the U.S. Naval Research Laboratory (NRL), the EPA Environmental Technology Verification (ETV) program Generic Protocol for the Verification of Ballast Water Treatment Technology (ETV Protocol), version 5.1, was published in 2010. This protocol applies solely to the land-based verification testing of BWMS, and another protocol is in preparation to provide guidance on shipboard testing. Likewise, this report pertains only to land-based testing.

The goals of this Intercomparison Project were to determine to what degree two independent, domestic ballast water test facilities (TFs) could comparably evaluate the performance of the same BWMS and to identify and quantify the variability between the facilities. Specifically, the TFs' methods and results would be used to assess the degree to which the facilities could follow the ETV Protocol, and if possible, the IMO Guidelines for Approval of Ballast Water Management Systems (G8). To meet these goals, the USCG Research and Development Center (USCG-RDC) contracted NRL to: purchase a commercial-off-the shelf BWMS; modify a Test Facility Questionnaire (TFQ) to assess candidate facilities' qualifications and readiness to conduct testing; solicit TF participation; recommend two TFs to participate in the evaluation; conduct three, valid biological efficacy (BE) tests following the ETV Protocol at each of the two TFs, yielding two verification reports; and compare the performance of the two TFs in testing in accordance with the ETV Protocol. All objectives were successfully met. The project commenced in July 2009, and final verification reports were received from the TFs in May 2012.

The two facilities selected were the Golden Bear Facility (GBF) and the Great Ships Initiative (GSI). Due to scheduling factors, the first TF to conduct testing was GBF, located in Vallejo, California aboard the California Maritime Academy's (CMA) training ship, the T/S *Golden Bear*. The tests were conducted according to the ETV Protocol for land-based facilities using the ship and its ballast systems as the TF while the vessel was at dock. The vessel was moored in the Carquinez Strait, where the salinity of the brackish water typically ranges from 10-20 practical salinity units (psu). The second TF to conduct testing, GSI, is a land-based TF located in Superior, Wisconsin along the shore of the Duluth-Superior Harbor in Lake Superior. The salinity at this freshwater site typically ranges from 0-1 psu. Thus, the BWMS was tested under two salinity regimes, at locations with differing biological communities and water quality parameters.

Each facility was contracted to conduct three valid BE tests according to the ETV Protocol draft version 4.2 (which was the most recent version of the protocol as the Intercomparison Project began). For the purposes

of this project, version 4.2 was indistinguishable from the final, published version (5.1). RDC provided the technical oversight of the testing as the Technical Authority (TA). The TA reviewed the technical objectives, test execution at both TFs, and test and analysis results for adequacy, relevancy, and usefulness. The TA reviewed and approved, with Executive Steering Advisory Committee (ESAC) input, the technical aspects of major milestones and deliverables. The NRL served as the Verification Organization (VO) and, as such, was responsible for supervising the development of all documents related to testing, witnessing the BE tests, and providing guidance for and approving the test plan documents and final verification reports. In addition, NRL's role was to analyze trends and variability in the BE test data and identify TFs' deficiencies or departures from the ETV Protocol. NSF International (NSF), an EPA ETV partner in operating the ETV Water Quality Protection Center, served as the Audit Organization (AO). In doing so, they assisted in document reviews, conducted audits of both facilities, witnessed BE tests, and provided feedback to the facilities. Both the VO and AO participated in the project's ESAC. This committee, established by and including members from the USCG-RDC, reviewed and approved recommendations proposed by the VO and AO during the course of the Intercomparison work. The ESAC also included representatives from the USCG Environmental Standards Division, the U.S. Maritime Administration, and the EPA.

Prior to starting the BE testing at the two facilities, a commercial-off-the-shelf BWMS was procured. It was selected according to the following specifications: it was a complete, commercially available BWMS, it was capable of treating ballast water without using "active substances" (e.g., chemicals or biocides), and its cost was compatible with the project budget. Ideally, the system had undergone land-based and shipboard verification testing and had received a Type Approval Certificate according to the IMO G8 guidelines. Requests for quotations were submitted to seven BWMS vendors; five responded. The selected BWMS filtered particles and organisms nominally ≥ 50 µm and treated smaller organisms with ultraviolet light and had received Type Approval from a party to the IMO. Per the specifications, the BWMS was configured inside a standard 20' (6.1 m) shipping container and could easily be transported between the two TFs.

The ETV Protocol specifies the concentrations of living organisms and multiple water quality parameters in the "challenge water" (i.e., uptake water) used in BWMS testing. As the project began, neither TF was able to meet all challenge water conditions with their existing infrastructure. Due to time and budgetary constraints, and solely for the purposes of this research project, departures from specifications (DFSs) were granted to both TFs to lower the requirements in the ETV Protocol for the concentrations of living organisms as well as dissolved and particulate matter. At GBF, a concentration of living organisms ≥ 50 µm of 10,000 m^{-3} was acceptable (the ETV Protocol requires 100,000 organisms m^{-3}), and concentrations of 3 mg L^{-1} of dissolved organic matter, 2 mg L^{-1} of particulate organic matter, and 20 mg L^{-1} of total suspended solids were acceptable (the ETV Protocol requires 6 mg L^{-1}, 4 mg L^{-1}, and 24 mg L^{-1}, respectively). At GSI, organisms in the ≥ 10 µm and < 50 µm size class were measured differently from the ETV Protocol. That is, for this project, it was acceptable to include in this size class living entities (individual cells or groups of cells) measuring at least 10 µm on *any* axis as long as individual cells measured < 50 µm in minimum dimension (i.e., the maximum dimension on the smallest axis). The ETV Protocol requires the maximum dimension on the *smallest* axis is at least 10 µm and less than 50 µm. This DFS allowed more organisms to be included in the size class than would be included in an ETV test. In the GSI verification report, organisms in this size class in uptake water were reported following both measurement approaches (the DFS and the ETV Protocol), whereas discharge measurements were reported only according to the ETV Protocol.

Without these reductions, neither TF met all water requirements in *any* test. Both TFs are situated in waters that can have very high concentrations of some of the challenge water constituents. If these TFs were unable to meet all the challenge water requirements, other facilities in the U.S. and abroad likely face similar issues. Thus, the challenge water requirements in the ETV Protocol, the means to augment ambient water to meet them, or both should be revisited.

Three valid BE tests were conducted at each TF (i.e., each test was completed and biological and engineering parameters were accepted by the VO). The requirements in the ETV Protocol for the concentration of living organisms in the control discharge water were met in 16 of 18 cases (3 size classes x 3 tests per facility x 2 facilities). In all BE tests at both TFs, the BWMS reduced the number of living organisms in treated water relative to uptake water (Table ES-1). Overall, the largest reductions were in the living organisms ≥ 50 µm: in the three tests at GBF, mean concentrations were reduced from 18,900-86,300 m^{-3} in challenge water to 127-284 m^{-3} in treatment discharge water; at GSI, mean concentrations were reduced from 341,000-396,000 m^{-3} in challenge water to 25,400-77,600 m^{-3} in treatment discharge water. However, neither TF obtained biological samples that approached their limit of detection for the two largest size classes, and in some instances, needed to modify protocols to accommodate higher-than-expected discharge densities.

Table ES-1. Summary of results from biological efficacy testing at both test facilities.

Sample source	GBF mean (SD) live counts			GSI mean (SD) live counts		
	BE 1 11 JAN 2011	**BE 2 18 JAN 2011**	**BE 3 25 JAN 2011**	**BE 1 12 JUL 2011**	**BE 2 25 JUL 2011**	**BE 3 09 AUG 2011**
≥ 50 µm size class (determined using light microscopy, m^{-3})						
Uptake[A]	86,300 (11,000)	18,900 (2,150)	37,100 (4,050)	341,000	353,000	396,000
Control[A] Discharge	55,200 (7,920)	30,100 (3,530)	50,000 (4,610)	413,000	593,000	559,000
Treatment Discharge[B]	127 (31)	217 (62)	284 (72)	77,600 (13,400)	25,400 (5,810)	48,100 (3,890)
≥ 10 and < 50 µm size class (determined using epifluorescent microscopy, mL^{-1})						
Uptake[C]	320 (65.5)	1990 (284)	757 (97.0)	605	572	560
Control Discharge	703 (185)	958 (295)	413 (87.1)	448	389	674
Treatment Discharge	9 (2)	7 (0.6)	4 (2)	301	197	95
< 10 µm size class (determined using HPC for GBF, IDEXX® kit for GSI, mL^{-1})						
Uptake	1700 (1,300)	2350 (1,390)	994 (466)	10,000 (4,000)	190,000 (30,000)	86,700 (30,600)
Control Discharge	409 (356)	259 (141)	511 (196)	7,300 (460)	40,000 (20,000)	113,000 (16,000)
Treatment Discharge	4 (5)	150 (120)	8 (10)	10,400 (2200)	15,800 (3,400)	19,900 (8,700)

[A]At GSI, organisms collected for this size class in the uptake and control discharge samples were *typically* ≥ 50 µm in minimum dimension (Great Ships Initiative, 2012, TQAP Appendix 6, SOP GSI/SOP/LB/RA/SA/2) and may have included some organisms < 50 µm in minimum dimension. [B]At GSI, all living organisms in treatment discharge samples are usually measured, but in this study, the density of organisms was higher than typically found, so a subsample of organisms was measured, and the proportion of organisms < 50 µm in the sample was extrapolated to

exclude organisms < 50 µm in the final, reported densities. [C]GSI values in this size class without SDs represent a single, composited sample; GSI values in this table were from organisms measured according to the ETV Protocol (i.e., the maximum dimension on the smallest axis was ≥ 10 µm and < 50 µm). BE = biological efficacy, cfu = colony forming unit, DOC = dissolved organic carbon, GBF = Golden Bear Facility, GSI = Great Ships Initiative, HPC= heterotrophic plate counts for culturable, aerobic, heterotrophic bacteria, SD = 1 standard deviation, and TQAP = test quality assurance plan. The values in this table were copied from the test facilities' summary tables (Table 3 and Table 5 in this report).

Variability was assessed within a facility and between facilities. The results showed that challenge water concentrations for the two largest size classes of living organisms within each facility varied as much as a factor of six over the three BE tests. This result is not surprising, given the natural variation in biological communities over a period of three to four weeks. The bacteria data were more variable, which is also not unexpected, given that their concentrations can change on an hourly basis. The biological concentrations for the two largest size classes within control discharge or treatment discharge varied at most by a factor of three across BE tests within a facility, indicating the facilities are capable of conducting repeatable tests. When comparing the two facilities, the concentrations of living organisms in the three size classes collected in uptake, control discharge, and treatment discharge water were often significantly different. Furthermore, in treated water, the final concentrations of organisms in the two largest size classes differed within each TF, suggesting the efficacy of the BWMS differed according to the group of organisms evaluated (e.g., at GBF, in the ≥ 50 µm size class, 127-284 organisms m^{-3} were present, and in the ≥ 10 µm and < 50 µm size class, an average of less than 10 organisms mL^{-1} were present). The data from this study illustrate the importance of determining the efficacy of a BWMS using different size classes of organisms and the significance of conducting land-based validation testing at multiple salinities, which will contain different communities and thus may challenge BWMSs in different ways.

A goal of this project was to assess the ability of domestic TFs to conduct validation testing following the ETV Protocol. The protocol is exceptionally complicated relative to other protocols in the ETV program, owing to the multidisciplinary nature of the testing, the large volumes of water required to mimic shipboard conditions, and the sparse populations of organisms in treated water. The two facilities chosen for this effort had varying experience in treatment and treatment system testing: one had only recently performed a first (shipboard) test according to the IMO G8 guidelines; the other had been involved for some time in bench scale and land-based system testing, as well as testing protocol development. Each facility complied with many of the requirements of the ETV Protocol, although in some aspects, they employed different approaches and processes. The ETV Protocol calls for independent auditing, and facility audits were performed for both TFs by NSF, who regularly performs this activity for ETV programs. Findings and recommendations from the audit process will be reported separately by NSF (in preparation). Both TFs benefited from developing procedures needed to run the ETV Protocol, and they are committed to continued improvement. It is the authors' collective opinion that the ETV Protocol can be successfully implemented and executed by domestic TFs, given that the ability of TFs to meet the challenge water conditions is addressed, either by TFs augmenting uptake water or changing the requirements in the ETV Protocol. It should be noted that this research project was not an ETV test, and furthermore, the information presented here should not be considered ETV data. Importantly, the results from this project cannot be compared to results for this BWMS tested elsewhere–without the detailed methods and procedures used in other testing efforts, as well as the output from quality assurance and quality control (QA/QC) procedures, a valid comparison is impossible.

Recommendations

From the goals of this project, it follows that ways to reduce variability and improve consistency among TFs should be suggested. According to the major findings and observations of the TFs, VO, and the AO, the following actions are recommended:

- Re-visit the challenge water conditions outlined in the ETV Protocol.
- Provide additional guidance within the ETV Protocol for the verification report. A sample report, template, or series of checklists would guide the TFs and assist VOs and regulators by standardizing the level of detail, terminology, output from data quality indicators (DQIs), and data presentation in reports. Furthermore, it would allow easy comparison among multiple verification reports, which would be useful to stakeholders and regulators.
- Expand the section in the ETV Protocol describing QA/QC, e.g., include recommendations about the application of DQIs, internal and external audits, management reviews, document control (e.g., locked spreadsheet calculations), and the need to have staff members solely devoted to QA/QC.
- Provide statistical guidance on subsampling organisms in uptake and control discharge samples (the ETV Protocol currently discusses at length the statistical approach of sampling rare populations in treated water, but it does not address the statistical approach for other populations).
- Provide guidance for a TF to assess the overall error rate for their facilities. A methodology for test facilities to quantify the error associated with their measurements would allow stakeholders and regulators the opportunity to assess the overall limit of detection at a test facility.

(This page intentionally left blank.)

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

(This page intentionally left blank.)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

ATP	Adenosine Triphosphate
ANOVA	Analysis of Variance
ANS	Aquatic Nuisance Species
AO	Audit Organization
BAA	Broad Agency Announcement
BE	Biological Efficacy
BOT	Ballast Order Telegraph
BWMS	Ballast Water Management System
BWTS	Ballast Water Treatment System
cfu	Colony Forming Unit
CMA	California Maritime Academy
CV	Coefficient of Variation
DB	Database
DFS	Departure from Specification
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter
DP	Differential Pressure
DQI	Data Quality Indicator
EPA	Environmental Protection Agency
ESAC	Executive Steering and Advisory Committee
ETV	Environmental Technology Verification
FBO	Federal Business Opportunities
g	Gram
G8	International Maritime Organization Guidelines for Approval of Ballast Water Management Systems (G8)
gal	U.S. Gallons
GBF	Golden Bear Facility
GFE	Government Furnished Equipment
GF/F	Glass Fiber Filter
gpm	U.S. Gallons per Minute
GSI	Great Ships Initiative
h	Hour
HMI	Human-Machine Interface
HPC	Heterotrophic Plate Count
HZ	Hertz
IMAC	Integrated Monitoring and Control
IMO	International Maritime Organization

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS (Continued)

ISO	International Standards Organization
kPa	Kilopascals
LoD	Limit of Detection
LSRI	Lake Superior Research Institute
m^3	Cubic Meter
$m^3\ hr^{-1}$	Cubic Meters per Hour
MARAD	U.S. Maritime Administration
MERC	Maritime Environmental Resource Center
mg	Milligram
$mg\ L^{-1}$	Milligram per Liter
min	Minute
MIPR	Military Interdepartmental Purchase Request
mL	Milliliter
MLML	Moss Landing Marine Laboratory
mm	Millimeter
MM	Mineral Matter
MT	Metric Tons
NEMWI	Northeast Midwest Institute
nm	Nanometer
NRL	United States Naval Research Laboratory
NRRI	Natural Resources Research Institute
NSF	NSF International
O&M	Operations and Maintenance
OIS	Organism Injection System
P&ID	Plant and Instrumentation Diagram
PNNL	Pacific Northwest National Laboratory
POC	Particulate Organic Carbon
POM	Particulate Organic Matter
ppm	Parts per Million
ppt	Parts per Thousand
PS	Percent Similarity
PSU	Practical Salinity Units
PVC	Polyvinyl Chloride
QA	Quality Assurance
QAPP	Quality Assurance Project Plan
QA/QC	Quality Assurance Quality Control
QMP	Quality Management Plan
RFI	Request for Information

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS (Continued)

RFP	Request for Proposals
RPD	Relative Percent Difference
s	Second
SCADA	Supervisory Control and Data Acquisition
SD	Standard Deviation
SIS	Sediment Injection System
SOP	Standard Operating Procedure
STO	Standard Test Organism
TF	Test Facility
TFQ	Test Facility Questionnaire
TQAP	Test Quality Assurance Plan
T/S	Training Ship
TSS	Total Suspended Solids
URS	URS Water Institute
USCG	United States Coast Guard
USCG-RDC	United States Coast Guard Research and Development Center
UV	Ultraviolet
UWS	University of Wisconsin-Superior
VAC	Voltage Alternating Current
VDC	Volts Direct Current
VO	Verification Organization
WET	Whole Effluent Toxicity
WQPC	Water Quality Protection Center
°C	Degrees Celsius
µm	Micrometer
’	Foot
”	Inch

ACKNOWLEDGEMENTS

We appreciate the input from the stakeholders of the Executive Steering and Advisory Committee (ESAC) throughout all phases of the Intercomparison Project. While the information and content in this report are solely the responsibility of the authors, the project in general and the report specifically were greatly improved by discussions with and feedback from ESAC members. In particular, we thank the following people: Dr. Richard Everett from the USCG Environmental Standards Division; Mr. Tom Stevens from NSF International, as well as Dr. Robert Donofrio and Mr. Joe Terrell, who participated in the audits of the test facilities and discussions about the results; Mr. Ray Frederick from the Environmental Protection Agency; and Dr. Carolyn Junemann from the U.S. Maritime Administration. Finally, we thank the members of the testing teams at the Golden Bear Facility and Great Ships Initiative for their sincere efforts to complete the testing according to schedule, to deliver quality data, and to construct well-written and thorough verification reports.

1 INTRODUCTION

The testing of ballast water management systems (BWMSs) to verify their performance regarding the discharge of living organisms against numeric standards—such as those proposed by the U.S. Coast Guard (USCG; 2009) and promulgated by the USCG (USCG, 2012) or the International Maritime Organization (IMO, 2004) (Table 1)—is a relatively new activity. To support verification testing of BWMSs, the international community, through the IMO, established guidelines for land-based and shipboard verification testing (G8 guidelines; IMO, 2005). To address land-based testing in the U.S., the USCG and the U.S. Environmental Protection Agency (EPA), with the input of stakeholder groups, collaborated to develop the Environmental Technology Verification (ETV) program protocol, which describes test facility (TF) requirements and procedures for the land-based verification of BWMSs (EPA, 2010). This "Generic Protocol for the Verification of Ballast Water Treatment Technology" is hereafter referred to as the "ETV Protocol".

Table 1. Ballast water discharge standards promulgated by the U.S. Coast Guard and International Maritime Organization.

Living organisms ≥ 50 µm in minimum dimension[A]	**Living organisms ≥ 10 µm and < 50 µm in minimum dimension[B]**	**Toxigenic *Vibrio cholerae*[C]**	***Escherichia coli***	**Intestinal enterococci**
$<10\ m^{-3}$	$<10\ mL^{-1}$	< 1 cfu $100\ mL^{-1}$ [D]	< 250 cfu $100\ mL^{-1}$	< 100 cfu $100\ mL^{-1}$

[A]Nominally zooplankton. [B]Nominally protists. [C]Serotypes O1 and O139. [D]The International Maritime Organization's proposed standard allows this metric to be measured alternatively as < 1 colony forming unit per gram of zooplankton wet weight. cfu = colony forming unit. Note that the shading in this and subsequent tables is to assist the reader.

Since the G8 guidelines and ETV Protocol became available, multiple TFs in the U.S. and abroad have begun construction and, in several cases, begun operation. To date, these facilities have primarily been independently developed and operated, although effort is underway to increase communication among facilities operating in different countries through the IMO's GloBallast Program. In the absence of standardized or scientifically validated and accepted methodologies, the TFs have provided fertile ground for the development of novel approaches to many of the challenges associated with full-scale testing of BWMSs including (1) sampling systems and methods, (2) biological enumeration and classification of organisms, (3) selection and use of surrogate or test organisms, and (4) modifications to the current protocols. Furthermore, the TFs are located in different geographic areas and necessarily operate under different environmental conditions. Thus, inherent variability among facilities—both with respect to physical configuration and biological and physical characteristics of challenge waters—will exist, and consequent variations in test results are expected.

Reducing variability and increasing comparability among TFs is necessary so the test data generated will provide confidence to regulators as they evaluate BWMS data packages for Type Approval. It will also allow ship owners to better examine performance tradeoffs as they choose BWMSs for their vessels. Furthermore, variations in test conditions across facilities must be accounted for and minimized to the extent possible because they could (1) create an unfair advantage among the TFs, with manufacturers preferring a facility that gives the 'best' results, or (2) create an unfair advantage among manufacturers, who might be able to minimize production costs by designing equipment for a specific facility where test results would be favorable, or (3) both.

The ETV Protocol recommends testing at two salinities, which would likely entail testing at two or more facilities. It is desirable to understand the degree to which these variations have or will affect the consistency of TFs' results. Moreover, critical techniques or processes that should be specified or normalized among TFs must be identified to ensure that the results of TFs using similar protocols will result in comparable test results. To address these issues, the USCG Research and Development Center (USCG-RDC) contracted the U.S. Naval Research Laboratory in Key West, Florida (NRL) to conduct an "Intercomparison Project".

Applications for participation were solicited from U.S. TFs that had responded to a prior request for information (RFI) regarding the availability of TFs (USCG-RDC, 2009). The main part of the application was a questionnaire that assessed TFs' capabilities in four areas: physical capability and operational conformance to the ETV Protocol, personnel to conduct the tests, ability to meet the test schedule, and costs to support all phases of the program. Following receipt of applications, two facilities were selected to participate. Concurrently, a commercial BWMS was selected and purchased. Each TF developed all necessary documents for testing (e.g., a test quality assurance plan, TQAP), and they conducted three biological efficacy (BE) tests. Operations and maintenance (O&M) testing, a required part of the ETV Protocol, is rather straightforward and does not entail new testing and analysis procedures, which were the focus of this project's effort. Thus, O&M testing was outside the scope of this work and was not conducted. The BE results of the Intercomparison Project provided an objective view of TFs' differences in capabilities, methods, and results; illustrated the major causes of variability; and allowed the facilities to recommend ways to improve consistency among facilities and increase testing efficiency.

This report first discusses the project goals, organization, and selection of the BWMS. Next, the process used to identify two TFs is explained. Afterwards, aspects of the two TFs are considered separately: their TQAPs, the commissioning process, and BE testing. Finally, protocol execution and results from the TFs are compared, and recommendations to improve the ETV Protocol are presented.

2 PROJECT OVERVIEW

2.1 Goals and Objectives

The goals of this effort were to objectively evaluate BWMS verification testing at two TFs and quantify the variability between facilities. In this manner, stakeholders would be informed about the practicability of TFs conducting comparable verification tests in accordance with the ETV Protocol, and if possible, the degree to which TFs could conduct testing according to the IMO G8 Guidelines for Type Approval of BWMS. The resulting analyses were intended to provide an objective view regarding the differences between TF capabilities, methods, and results. These analyses examined comparability within and between TFs and recommended methods to improve consistency, comparability, and accuracy among TFs using the same procedures.

The objectives to accomplish these goals were to procure a BWMS, identify two domestic TFs deemed likely to be able to conduct the tests, advise the facilities on the preparation of all necessary documentation (test plans, etc.), conduct three BE tests following the ETV Protocol, and compare the TFs' methods and results. Notably, this project was initiated before the final version of the EPA Protocol was published (v. 5.1); thus, version 4.2 was used by the TFs. The two versions were indistinguishable for purposes of the Intercomparison Project because the differences were editorial rather than substantive.

Importantly, the Intercomparison Project was designed to determine the practicability of conducting verification tests according to specific, key portions of the ETV Protocol regarding biological efficacy. The results generated through testing were intended neither to be used for purposes of Type Approval by the USCG nor to endorse any vendor's BWMS. Instead, this effort represents a research project on the ability of two TFs to conduct comparable tests in accordance with a common protocol.

2.2 Project Management and Organization

The USCG-RDC provided the technical oversight of the project as the Technical Authority (TA). The role of the TA was to 1) review the technical objectives, test execution at both TFs, and test and analysis results for adequacy, relevancy, and usefulness; 2) review and approve the technical aspects of major milestones and deliverables; and 3) establish and chair the Executive Steering and Advisory Committee (ESAC). ESAC's role was to review and provide input to the recommendations made by the Verification Organization (VO) and Audit Organization (AO), review major deliverables, and make recommendations to the TA. In addition to representatives from USCG-RDC, ESAC consisted of representatives from the NRL, USCG Environmental Standards Division (CG-OES-3), EPA Office of Research and Development ETV Water Quality Protection Center (WQPC), U.S. Maritime Administration (MARAD), and NSF International (NSF).

The VO was NRL, whose role was to oversee the project, including 1) develop selection criteria for TFs and the BWMS used in testing, 2) recommend the TFs and BWMS to ESAC for their approval, 3) procure, handle, and transport the BWMS to and from the TFs, 4) oversee development of, and approve, the test plan and other documentation for both TFs, 5) observe testing activities conducted by the TFs, 6) provide guidance on, and approval of, the TFs' verification reports, 7) coordinate and report to ESAC, 8) coordinate with the AO and incorporate the AO's findings into the final report, 9) conduct comparative analyses between the TFs based on the verification reports prepared by the TFs and the observations of the VO during planning and testing by the TFs, and 10) prepare a comprehensive report addressing the proposed objectives. The AO was NSF, who as EPA's formal partner for the operation of the ETV Water Quality Protection Center, has a unique understanding of the EPA's process and expertise in reviewing Quality Assurance Project Plans (QAPPs) and conducting audits of facilities performing ETV testing. In this project, NSF's role was to 1) oversee development of and provide feedback on the test plan and other documentation for both TFs, 2) audit each TF in advance of testing, 3) observe testing and validate that the procedures and data quality were in accordance with the QAPP, 4) report to ESAC, and 5) provide a report to USCG-RDC.

2.3 Government Furnished Equipment Selection

The commercially available BWMS selected for this project was a complete system suitable for shipboard installation and use. The intent was to use the equipment at the TFs as if it were being used in a real-world situation, that is, without more assistance or influence from the manufacturer of the equipment than would normally occur when a unit was purchased. The participants in the project referred to the BWMS as "government-furnished equipment" (GFE), as this generic description separated the manufacturer of the equipment from the overall goals of this research project, i.e., to evaluate the capabilities of TFs to execute the ETV Protocol in a comparable manner and to quantify variability between facilities.

The criteria for choosing the BWMS included the following items: (1) the system had to be a complete, commercially ready BWMS, (2) the overall cost of the system had to be compatible with the project budget,

and (3) the system needed to be capable of treating ballast water without the use of "active substances" (e.g., chemicals or biocides). The use of active substances would have introduced additional complications to the project: permits for discharges from TFs might have been required, and the relevant federal, state, and local regulations would have had to be reviewed. Thus, a BWMS without active substances was preferable in this instance. Ideally, to increase the likelihood that the BWMS would operate consistently throughout the tests at both TFs, a system was desired that had undergone land-based and shipboard verification testing and had been granted a Type Approval certificate by a foreign administration in accordance with the IMO ballast water management convention (IMO, 2004) and following the IMO G8 guidelines. These selection criteria ensured the system was complete, would be easily configurable at the participating TFs, and had already been demonstrated to operate consistently.

Quotations were requested from seven vendors of BWMSs who were identified through publicly available information (e.g., websites, trade show presentations, IMO submissions, etc.). Five responded. The following specifications were provided to the vendors:

- Capability to treat ballast water up to a flow rate of 250 m^3 h^{-1} at varying pressures.
- Ability to treat water over a range of salinities (salinity range of freshwater with <1 practical salinity units [psu] to marine water with 28-36 psu) and temperatures (4-35°C).
- Ability to operate in ambient air temperature between 0-35°C.
- Configuration to accept a 440VAC/3phase/60 HZ power supply that could be stepped down to 220VAC or 110 VAC.
- Capacity to use the compressed air supply (of at least 90 pounds per square inch [psi]) at the participating TFs.
- Contain piping to accept a supply and return water connection using 6" (15 cm) flanges.
- Capable of accepting and sending external automated control signals to start and stop the system; capable of providing alarms to the TFs' control systems.
- Flexibility to be configured inside a standard 20' International Standards Organization (ISO) shipping container for easy shipping between TFs.

After reviewing the quotations submitted from the five vendors, a BWMS using filtration plus ultraviolet (UV) light was selected. This BWMS treated ballast water on uptake using filtration and UV light and treated ballast water on discharge using UV light. Ballast water was filtered through openings with a nominal dimension of 50 µm, and the medium pressure UV-C lamps operated at a wavelength of 240-300 nm. This system had also been tested according to the IMO G8 guidelines, received Type Approval, and had been installed on commercial ships, which provided a measure of confidence in the system's maturity.

3 TEST FACILITY SELECTION, DEPARTURES FROM TESTING REQUIREMENTS, AND PROJECT TIMELINE

3.1 Overview

Based on the type and complexity of the information and services required for this project, quotations were requested from the four facilities that had responded to the USCG-RDC RFI. A fifth facility, the T/S *Golden Bear* (Golden Bear Facility, GBF) at the California Maritime Academy (CMA), was known by ESAC members to be developing the capabilities to test BWMSs but had not responded to the RFI. Based

on the existing knowledge, GBF was considered by the ESAC to be a potential candidate for testing under this project and was added to the original list of responders to the RFI.

3.2 Test Facility Questionnaire

After draft version 4.2 of the ETV Protocol was completed the VO had developed a tool for the USCG (the "Test Facility Questionnaire" [TFQ]) to assess a TF's ability to conform to the ETV Protocol. For the Intercomparison Project, the TFQ formed the basis for each TF to provide information on its capabilities to conduct ETV Protocol testing. The TFQ was modified to allow each facility to enter a proposed time line, identify physical improvements or facility developments that would be needed for compliance to ETV requirements, and list any anticipated issues or problems in testing to the ETV Protocol.

It was acknowledged that none of the facilities would likely be fully capable or compliant with the draft ETV Protocol, but to assess the amount of effort needed to achieve compliance with the protocol, details on each facility's capabilities, methodology, and staffing were needed. The TFQ provided each facility an opportunity to identify areas of the protocol with which they were and were not compliant and to supply documentation for additional details. This process allowed the reviewers to assess any requisite changes and verify that critical parts of the ETV Protocol would be addressable by at least some of the facilities. Using the same tool for all facilities to submit information also allowed a standard, objective rubric to be developed by the VO to compare facility capabilities.

All five facilities were provided with the updated TFQ, ETV Protocol version 4.2, and the scope of the proposed testing, and they were requested to submit a quotation for participation in the Intercomparison Project (this request for proposals [RFP] was dated 24 MAR 10). A sample TFQ that had been completed by the VO was also provided to the facilities to illustrate the type and level of detail of information needed. As the TFQ is extensive, it is not included in this report but it is available as a letter report (Lemieux et al., 2010b). The TFQ was designed so each TF was evaluated according to the following four general criteria:

1. Physical capability and conformance to the ETV Protocol
2. Timeline for planning, verification testing, and reporting for a BE test.
3. Available personnel, including their training and experience.
4. Costs of testing and operation as well as costs to modify or augment facility capabilities to adequately conduct testing as described and intended in the ETV Protocol.

To address the first criterion, the TFs were asked to supply several documents: a "Facility Technical Documentation Package", which included, at a minimum, a map(s) or diagram(s) showing the TF layout and dimensions, plant and instrumentation diagrams (P&IDs), and an equipment list(s) for pumps, sensors, valves, etc., and their relevant capacities and ranges; a Quality Management Plan (QMP); a Laboratory Quality Assurance Manual; the relevant Standard Operating Procedures (SOPs); an Environmental, Health and Safety (EHS) Plan; and a QAPP. This requested documentation, taken together, illustrated the ability of a facility to run standardized test protocols with appropriate scientific, quality, and management practices. In particular, the elements included in the Facility Technical Documentation Package of the TFQ allowed an assessment of the physical capabilities and test methodologies that correspond to requirements of the ETV Protocol. This section of the questionnaire requested specific details to assess the "What do you do?" and the "How do you do it?" of ballast water treatment testing in the following areas: test verification factors, challenge water conditions, physical construction of the facility, and data management. The remaining sections of the TFQ requested details on test timelines, personnel, and cost (criteria 2-4 in the list above).

3.3 Test Facilities Chosen for the Intercomparison Project

Based on gathered information and TFQ responses, the VO recommended to ESAC and the TA that the project proceed with the GBF and the Great Ships Initiative (GSI) as the Intercomparison facilities. ESAC concurred, and the TA approved the recommendation. The selection ensured that each facility was located in a biologically productive area, one estuary and one freshwater environment. Thus, the realized challenge conditions during testing were of natural and diverse communities. The work to be performed was verification testing of a government-furnished BWMS according to the ETV Protocol, including all associated reporting requirements. The verification reports would not be reviewed by the EPA (as per the ETV Protocol), but for the purposes of this research project, they would require review by the VO (NRL) and the AO (NSF), concurrence by ESAC, and approval by the TA (USCG-RDC). BE testing consisted of three successful tests at a single salinity at each TF. Operational and maintenance testing of 10,000 m^3 of treated water was excluded from the testing.

Golden Bear Facility: GBF agreed to validate the proposed sampling of three replicate time-averaged sample volumes of 3 m^3 and fully document methods and statistical justification per the ETV Protocol in the GBF TQAP. GBF would use their best effort using existing facility equipment to maximize biological densities in challenge water and was allowed a departure from specification (DFS) to three of the ETV challenge requirements (see Section 3.6 below "Departures from ETV Challenge Water Testing Requirements"). Testing was scheduled to occur in January 2011.

Great Ships Initiative: The GSI facility agreed to validate any changes made to facility infrastructure, methods, and procedures over the 2010 testing season prior to commencement of Intercomparison tests. Furthermore, GSI was to document fully the sampling methods and statistical justification per the ETV Protocol in the TQAP. The facility would use existing equipment (which included augmentation systems for organisms ≥ 10 µm and < 50 µm as well as two water-quality parameters, particulate organic carbon [POC] and mineral matter [MM]). A DFS regarding the sizing of organisms in the ≥ 10 µm and < 50 µm size class was allowed (see Section 3.6 "Departures from ETV Challenge Water Testing Requirements"). Testing was scheduled to occur in spring of 2011 as soon as plankton densities reached suitable levels for ETV testing, and Intercomparison testing would have priority over other full-scale, land-based testing opportunities that might arise.

3.4 Overview of ETV Testing Requirements

The goal of land-based verification testing under the ETV Protocol is to verify the performance of BWMSs under conditions that are (1) tightly controlled and (2) rigorous, that is, representative of demanding conditions that might be encountered by a ship's BWMS. To address the rigor of the testing conditions, a set of challenge water conditions is specified in the ETV Protocol (Table 2). They address physicochemical properties, e.g., the concentration of dissolved organic matter (DOM), as well as biological conditions, e.g., the concentration of organisms in the groups stipulated by the ETV Protocol. This matrix aims to provide testing conditions representative of the water found in the world's ports that can present a challenge to BWMSs. Under these conditions, BE tests are conducted to determine the system's ability to "kill, remove, or inactivate organisms" after treatment and a hold time of at least one day (EPA, 2010). Furthermore, verification testing of a given BWMS requires at least three BE tests at each of two or more of the specified salinity ranges (for the Intercomparison Project, each TF conducted three BE tests in one salinity range).

Table 2. Requirements from the ETV Protocol (shown in the darkly shaded rows) and departures from specifications for the test facilities.

Protocol or TF Water type	Requirements for testing and departures from specifications granted to test facilities			
	Living organisms ≥ 50 µm Diversity req.	**Living organisms < 50 µm and ≥ 10 µm[A] Diversity req.**	**Living organisms < 10 µm[B]**	**DOM, POM, TSS Temperature range Salinity range**
ETV Challenge water	≥ 100,000 m^{-3} ≥ 5 species from ≥ 3 different phyla	≥ 1,000 mL^{-1} ≥ 5 species from ≥ 3 different phyla	1,000 mL^{-1} culturable, aerobic heterotrophic bacteria	DOM: 6 mg L^{-1} (as DOC) POM: 4 mg L^{-1} (as POC) MM: 20 mg L^{-1} TSS: 24 mg L^{-1} (= POM + MM) T = 4 – 35 ºC Salinity: < 1, 10 – 20, 28 – 36 psu
GBF Challenge water	≥ 10,000 m^{-3}	Same[C]	Same[C]	DOM: 3 mg L^{-1} (as DOC) POM: 2 mg L^{-1} (as POC) MM: 18 mg L^{-1} TSS: 20 mg L^{-1} (= POM + MM) Same Salinity = 10 – 20 psu
GSI Challenge water	Same[C]	Any entity with any dimension ≥ 10 µm and individual cells < 50 µm is acceptable[E]	Same[C]	Same[C] Same[C] Salinity = < 1
ETV Control tank discharge[D]	≥ 100 m^{-3}	≥ 100 mL^{-1}	≥ 500 mL^{-1}	NR

[A]Size determined by "maximum dimension on the smallest axis" (EPA, 2010). [B]The concentrations of toxigenic *Vibrio cholerae*, *Escherichia coli*, and intestinal enterococci are excluded because there is no requirement for their concentrations in the ETV Protocol; also, there are no diversity requirements for any of the bacteria. [C]"Same" indicates the requirement(s) for the TF is the same as the ETV requirement(s). [D]No deviations from specifications were needed for control tank discharge requirements. [E]Results meeting ETV Protocol requirements were also to be presented in the verification report; an 'entity' is an individual cell or a group of cells. DOC = dissolved organic carbon, DOM = dissolved organic matter, ETV = Environmental Technology Verification program, Golden Bear Facility = GBF, Great Ships Initiative = GSI, MM = mineral matter, NR = no requirement, POC = particulate organic carbon, POM = particulate organic matter, psu = practical salinity units, req. = requirement, T = temperature, and TF = test facility.

Testing under the ETV Protocol also carries requirements for infrastructure and engineering parameters: control and treatment tank capacity and water volumes used in testing must be at least 200 m^3 (~52,800 gal), water flow rates must be at least 200 $m^3 h^{-1}$ (880 gal min^{-1}), and other requirements (as outlined in the ETV Protocol must be met.

Full ETV BWMS testing is divided into three phases: (1) commissioning of the BWMS after it is installed at the TF, (2) BE testing, and (3) O&M testing (in which 10,000 m^3 is treated by the BWMS over the course of the three BE tests, entailing approximately 50 h of testing). Operations and maintenance testing was outside the scope of this project, but at each TF, the commissioning and BE testing phases were conducted.

3.5 Departures from ETV Challenge Water Testing Requirements

As previously noted, exceptions or DFSs were granted to each TF in advance of testing for specific challenge water requirements. Although both facilities were located in productive water bodies, reflective of challenge conditions encountered in estuarine and freshwater environments, the exceptions were necessary to meet the testing schedule and timeline for the project, given the pressing need to complete the work and associated reports to inform USCG-RDC and other federal partners of the outcome and results. The details prompting these exceptions are provided below for each TF.

Data collected previously by GBF regarding the concentration of ambient organisms and suspended material in the water from Carquinez Strait showed that the concentration of organisms in the ≥ 50 µm size class as well as concentrations of dissolved organic carbon (DOC) and POC would be unlikely to meet the challenge water conditions during all tests. At the commencement of the project, GBF possessed neither the infrastructure nor the validated procedures to augment ambient water with additional organisms or suspended material. Furthermore, given the timeline and budget of the project, it was not practicable to update the facility as well as validate and implement new procedures. Nonetheless, it was important to test the BWMS under challenging conditions. As a compromise, DFSs were allowed to reduce the requirement for organisms ≥ 50 µm by ten-fold and to reduce the requirements for DOC and POC by two-fold (Table 2).

GSI had validated the infrastructure and procedures to augment challenge water with POC and MM. Additionally, GSI routinely used a procedure they had validated to increase the number of organisms in the ≥ 10 µm and < 50 µm size class: a plankton net was slowly towed behind a motorboat, collected organisms were transferred to and held in an outdoor, uncovered, aerated holding pond prior to injection into the uptake water. The size of these ambient organisms, however, was typically < 10 µm as measured according to the ETV Protocol, defined as the "maximum dimension on the smallest axis" (EPA, 2010). If, however, GSI measured entities (i.e., individual cells or groups of cells) to include those with *any dimension* equal to or greater than 10 µm and individual cells less than 50 µm in minimum dimension, the challenge water conditions could be met. GSI did not have in place the infrastructure to culture organisms to meet the size requirements as defined by the ETV Protocol, nor did the timeline or budget permit the construction of such infrastructure. Additionally, the slightly smaller size of entities having *any dimension* equal to or greater than 10 µm and less than 50 µm relative to the "maximum dimension on the smallest axis" seemed unlikely to affect the results from the BWMS chosen for the project. That is, the BWMS used a filter to remove particles and living organisms ≥ 50 µm and UV light to treat organisms < 50 µm, so including organisms slightly smaller than 10 µm would seemingly have little effect on the BWMS efficacy, assuming UV radiation would have a similar effect on the organisms. Thus, as a compromise, a DFS in the approach for measuring organisms ≥ 10 µm and < 50 µm in the uptake water was granted (Table 2; see the example in Figure 1). This DFS applied to validity criteria used for uptake populations only; for reference, the uptake

values according to sizing criteria in the ETV Protocol are also reported. Notably, all discharge populations are reported only in accordance with the sizing criteria in the ETV Protocol.

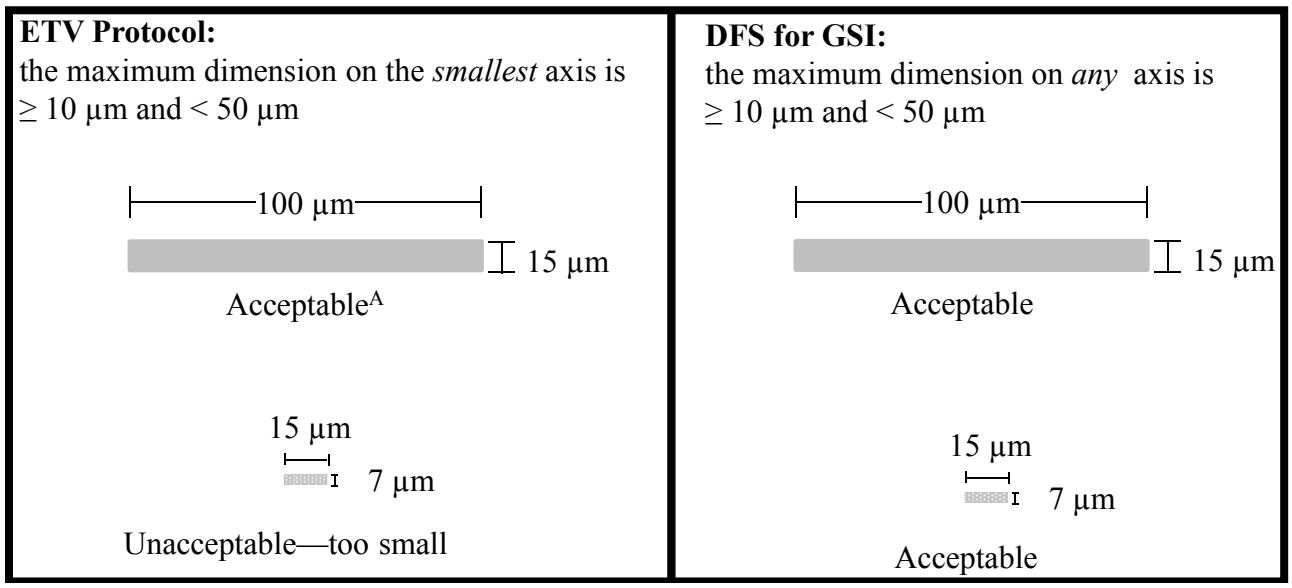

Figure 1. Example of approaches to measuring entities (i.e., individual cells or groups of cells) ≥ 10 µm and < 50 µm. [A]"Acceptable" indicates an organism with these measurements would be acceptable to count using the given measurement approach. DFS = departure from specification, ETV = Environmental Verification Protocol program, and GSI = Great Ships Initiative.

3.6 Intercomparison Timeline

The timelines in Figures 2-4 show when tasks commenced or were completed. From July 2009 to April 2010, the VO focused on development of TF requirements and selection of a BWMS and facilities (Figure 2). The contract to build the BWMS was issued in January 2010. In June 2010, GBF and GSI were selected as participating facilities.

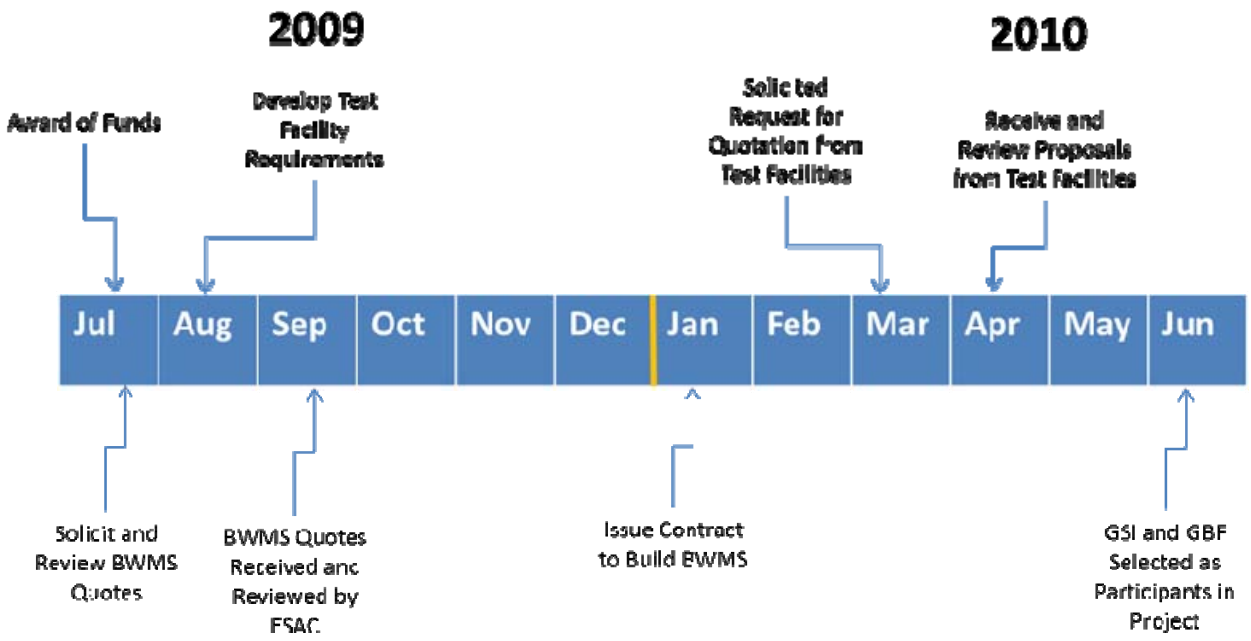

Figure 2. Timeline for the Intercomparison Project (July 2009 – June 2010).

As there was one BWMS to be used between the two TFs, and testing at GSI could not commence until spring or summer, GBF was selected as the first TF to conduct BE tests to meet the project schedule (Figure 3). GBF completed their TQAP, commissioned the GFE in early January 2011, and initiated BE tests soon thereafter. Concurrent to GBF testing, GSI drafted their TQAP and submitted the document in February 2011 to be ready for testing in early summer 2011. The commissioning of the BWMS at GSI was completed in June 2011.

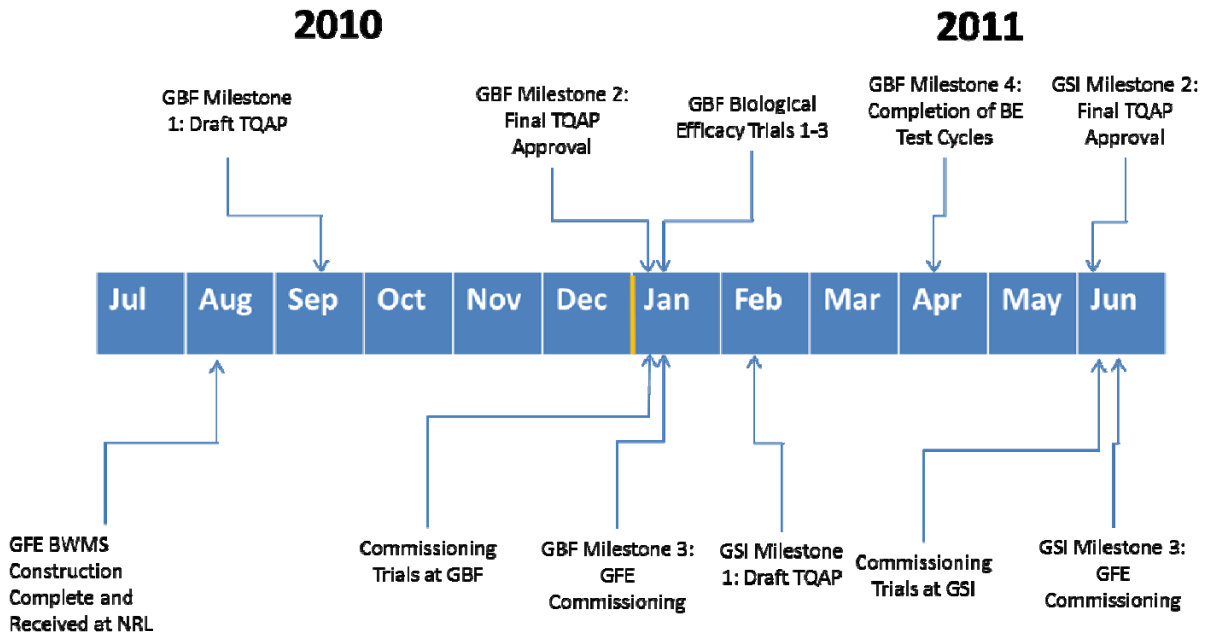

Figure 3. Timeline for the Intercomparison Project (July 2010 – June 2011).

BE tests were conducted at GSI in July 2011 (Figure 4). As GSI conducted tests, GBF analyzed data and assembled and submitted a first draft verification report in July 2011. GSI assembled and submitted their first draft verification report in November 2011. Both TFs were given written feedback on their drafts by the VO and the AO. The TFs, along with the VO, AO, ESAC, and TA, met in November 2011 in Warwick, RI to discuss results of the BE tests, lessons learned, and the ETV Protocol (specifically, changes the TFs recommended to Protocol). Over the winter, both TFs submitted drafts of their verification reports, and the final reports were submitted in May 2012.

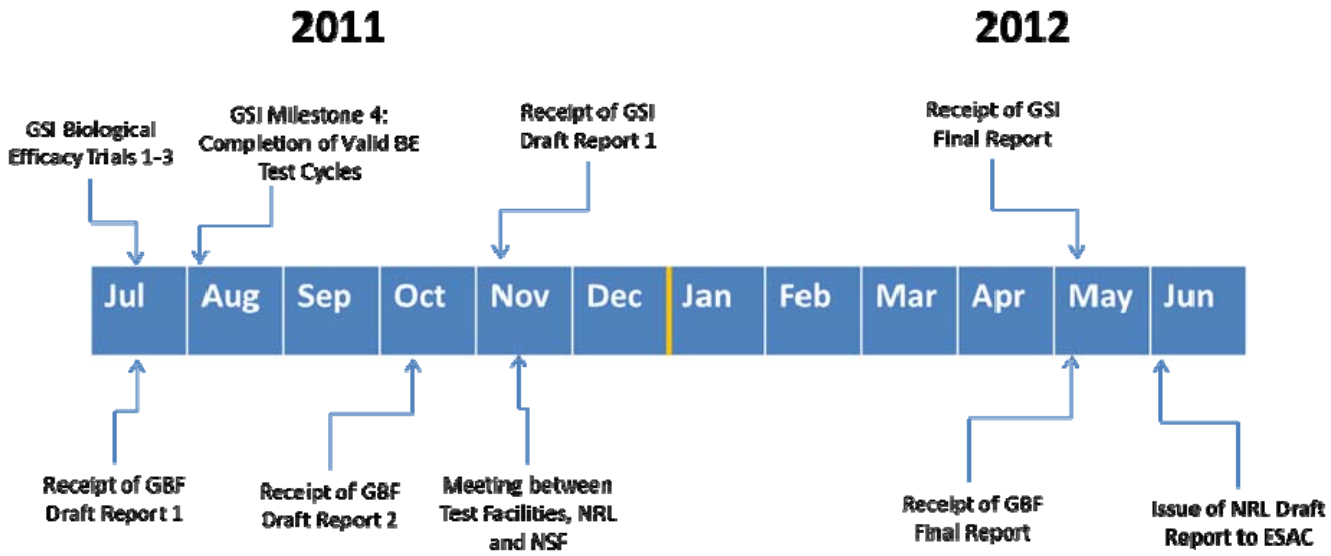

Figure 4. Timeline for the Intercomparison Project (July 2011 – June 2012).

4 GOLDEN BEAR FACILITY

4.1 Golden Bear Facility Test Quality Assurance Plan

A TQAP is defined in the ETV Protocol as "a written document that describes the procedures for conducting a test or study according to verification protocol requirements for the application of a particular ballast water treatment system at a particular site. At a minimum, the TQAP shall include detailed instructions for sample and data collection, sample handling and preservation, precision, accuracy, goals, and quality assurance and quality control requirements relevant to the particular site." (EPA, 2010). For this project, the TFs developed their TQAPs, had them reviewed by the VO and the AO, and received approval from the VO prior to BE testing.

4.1.1 Required Elements of the Test Quality Assurance Plan

The ETV Protocol lists the following requirements that should be included in a TF's TQAP (EPA, 2010):

- Title page/approval page with all project participants
- Table of contents
- Project description
- Project organization and personnel responsibilities
- Treatment system description
- Experimental design (including installation and start-up plan)

- Challenge water conditions and preparation (including TF SOPs)
- Sampling and analysis plan including sampling and analytical procedures
- Data management, analysis and reporting
- Environmental, health and safety plan
- References
- Appendices: Quality Assurance Project Plan
- Appendices: Vendor Operation and Maintenance Manual

In typical EPA ETV testing, the vendor funds an independent TF to generate verification data to validate the vendor's claims. For the Intercomparison Project, one BWMS was supplied to two TFs as GFE. Thus, no vendor claims were required or cited, nor did the vendor provide input into the TQAP development, other than to respond to queries from the TFs while writing their TQAPs. As this project was a research effort to examine the degree to which two independent TFs could implement and comply with the ETV Protocol, and most importantly, BE testing, no O&M (i.e., long-term performance) tests were required. The TFs, however, were required to generate documentation comparable to that of other ETV tests, according to ETV conventions and requirements defined by the EPA. To facilitate this, examples of required documentation from other types of ETV testing were provided to the TFs via the EPA's website.

4.1.2 Development Process

The first draft TQAP was received from GBF on 09 SEP 2010. Six documents were submitted: the Facility Description and Capabilities, Challenge Water Conditions, Engineering Safety and Health Plan, QMP, Project Plan, and QAPP. The first draft TQAP and other documents were reviewed by the VO and the AO, who provided comments to GBF and held a phone conference with GBF on 01 OCT 2010. After the teleconference, GBF submitted a revised TQAP on 26 DEC 2010. Following an additional round of review and revision, the final TQAP was received on 07 JAN 2011 and accepted by the VO. The process to develop the TQAP lasted several months, and the final version was ready the week before the first BE test occurred. All required elements of the TQAP (as above, in Section 4.1.1 "Required Elements of the Test Quality Assurance Plan") were included and addressed.

4.2 Golden Bear Facility Commissioning

The BWMS was successfully installed and commissioned aboard the T/S *Golden Bear* during the week of 16 DEC 2010. The installation of the system required piping and support modifications to the BWMS, as the inlet and outlet piping were located on the opposite ends of the container, which impeded installation aboard the vessel. Some relatively minor problems were encountered during commissioning, but GBF discussed all with the VO, and they were rectified by the end of the commissioning week. Prior to testing, a checklist was prepared by the VO and the relevant commissioning SOP reviewed. All relevant tasks were observed on-site by the VO and a representative from USCG-RDC.

The VO reviewed the BWMS installation and commissioning testing and verified that all commissioning activities planned by GBF were completed. During this trip, the VO also witnessed GBF personnel sampling organisms over tidal cycles to identify the point at which ambient waters would have the highest likelihood of achieving target challenge water conditions. Finally, as part of the commissioning, several operational changes were identified for GBF test procedures. These included lowering the initial control tank uptake flow from 200 $m^3 hr^{-1}$ to 185 $m^3 hr^{-1}$, performing in-test adjustment of control tank uptake flow to match average treatment uptake flow, modifying deballast operating procedures to allow for more

complete draining of the control and treatment tanks, and having all personnel remain outside of the container during test operations. These changes are further discussed below and were incorporated by GBF into the subsequent final revision of the TQAP. During commissioning, a concern was identified regarding an abrupt water hammer (a shock resulting from a pressure surge) observed during backflush operations of the BWMS; this behavior was worrisome for safety reasons and for the long-term durability of the system, and it is further discussed in the following sections. A technical representative from the BWMS manufacturer was on site for the commissioning; he provided training, verified equipment operation, and approved the installation of the equipment met manufacturer requirements.

4.2.1 Installation

The GBF facility is located on an operational training ship. To minimize disruption of training operations and facilitate short-term installation, the BWMS was installed as proposed in the TQAP, on the 01 deck, with connections to ship systems via interfaces installed specifically for this type of testing. Several modifications to the BWMS were completed by GBF and contracted engineers to accommodate the installation. The most significant modification was to allow water to enter and exit from the same end of the container; thus, the 6" (15.2 cm) diameter polyvinyl chloride (PVC) outlet line and the 4" (10.2 cm) diameter PVC drain line from the ballast management system were rotated 180°. This change also allowed the doors on one end of the container to remain closed while the BWMS was in operation. Additional pipe support hangers were installed on the ceiling of the container to hold up the outlet line.

The 440 VAC ship's power supply was wired to the main electrical disconnect by the GBF electrical personnel. GBF also plumbed the ship's compressed air supply to the BWMS air supply inlet with appropriate pipe and valve components. A controls communication cable was installed between the GBF Integrated Monitoring and Control system (IMAC) and the BWMS control panel, and GBF cut a hole through the container to insert the control communication cable. Slight damage that occurred during shipping was remedied by GBF prior to starting the BWMS: three filter fan covers from the UV system power supply had fallen to the floor, a piece of 0.5" (1.3 cm) conduit was broken, and a bolt on the 110 VAC transformer backing plate had come loose. All were repaired during installation.

4.2.2 Checkout and Operational Tests

Subsequent to the BWMS piping modifications, a 100 psi (~7 Bar) hydrostatic pressure test was performed (note that the ETV Protocol requires neither hydrostatic nor operational pressure testing). One leak was found and repaired; a second pressure test confirmed the repair was successful. Commissioning tests and manufacturer setup procedures were performed according to the GBF SOPs and the BWMS operation manual. These tests revealed one BWMS flow sensor had an incorrect calibration, and discrepancies were present in the Operations Manual that accompanied the BWMS. The technical representative from the manufacturer corrected the calibration issue on site and addressed the discrepancies through training with the GBF staff. Operational flow testing occurred for several hours using ambient water at a flow rate of 200 $m^3\ h^{-1}$, and several backflushes of the filtration system were observed at intervals of approximately 20 minutes. The UV system was verified as operational, the UV transmittance sensor worked correctly, and the UV lamp wiper system operated successfully. All applicable portions of the GBF installation checkout were successfully completed and documented.

The checkout test was followed by a "Shakedown Test" per GBF's SOP5 (Golden Bear Facility, 2012), in which the BWMS was "stressed" by operating the BWMS near its maximum flow rate of 250 $m^3\ h^{-1}$ in both "Ballast" and "Deballast" modes. The system was operated continuously and successfully for 4 h at a flow

rate of approximately 250 $m^3 h^{-1}$. The filtration system backflushed 23 times during the 4-h period (approximately every 10 min). The relatively high flush frequency was attributed to both the increase in flow rate and higher turbidity and sedimentation in the water flowing through the Carquinez Strait during this time compared to earlier test runs at GBF.

During this portion of commissioning, it became apparent that additional supports were required to stabilize the vertical section of the 6" (15.2 cm) outlet line within the container. When water flow resumed after stopping for the backflush cycle, an abrupt water hammer occurred as the BWMS outlet valve opened. To provide extra support in addition to the overhead brackets already installed, a steel L-channel was welded by GBF from floor to ceiling to mount pipe hangers that reinforced the vertical piping. This reinforcement significantly reduced the amount of deflection in subsequent tests. Even with this support in place, pressure pulses caused a noticeable flexing of the housing for the UV reactor. The pressure surge or water hammer was caused by rapid closing and opening of the outlet valve of the BWMS and the significant length of piping (approximately 15 m [50']) connecting the device to ship systems. The magnitude was sufficient to cause notable deflection of BWMS components even when reinforced. This was deemed a potential failure point and safety hazard, so the GBF SOPs were modified to require observers to remain outside the container during BE testing, and an emergency shutdown procedure was planned for the BE tests. An additional part of the Shakedown Test verified that the high temperature shutdown safety feature worked. When water flow to the BWMS was stopped, the system correctly detected a temperature rise in the UV reactor, operated the cooling valve, and subsequently shut down.

Although specified by the ETV Protocol, a complete rehearsal of BE test operations was not performed by GBF, as they believed that adequate subsets of operations were performed in conjunction with checkout testing. In addition, GBF had said the BWMS was comparable to other filtration systems that had been tested by GBF, so a full test was not necessary. The minimum requirement acceptable to the VO was a brief simultaneous uptake and discharge operation run without sampling. As a complete trial run was not performed, GBF provided a contingency test date in the event the first BE test was not valid.

The uptake operation was performed with the facility operating in sea-to-sea mode and flow rates set at 200 $m^3 h^{-1}$ to both the treatment and control tanks. Sea-to-sea mode here was defined as taking water in directly from the vessel's sea chest, then pumping it through the facility piping system, then bypassing the treatment and control tanks and discharging directly out of the vessel. The uptake operation was conducted with relatively clear challenge water (a qualitative, visual observation of the water flowing past the vessel). Thus, long periods (> 20 min) ensued between automatic backflush operations of the BWMS. While this source water did not appear to challenge the BWMS (given the relatively long time between backflush operations compared to those in the GBF BE tests, which are discussed below), this test demonstrated the need to balance the control flow to match the average of the treatment flow. The average treatment flow was lower than 200 $m^3 hr^{-1}$ because the water flow to the treatment tank completely stopped during backflush operations. To allow both control and treatment tanks to fill at the same rate, the control flow (which continued during backflush) needed to be lower than the target of 200 $m^3 h^{-1}$. Balancing the flow rates was accomplished using a manual pinch valve in the engine room. The GBF SOP was subsequently modified to reflect these changes. Another change following commissioning was to reduce the deballast flow rate from 200 $m^3 h^{-1}$ to 50 $m^3 h^{-1}$ over the last several minutes of a tank discharge operation to allow the tank (either control or treatment) to be more completely drained than if the flow rate was not reduced. Slowing the pumping rate allowed the tanks to be drained further before losing suction.

4.2.3 Other Tests during the Commissioning Period

To identify the optimal timing for challenge water uptake during the tidal cycle (per the GBF TQAP), GBF sampled ambient water to monitor both water chemistry and biological parameters. As mentioned above, GBF completed stressing cycles, setting the flow rate to 250 m^3 h^{-1} for 4 h, which was within the BWMS operating specifications but set at a flow rate higher than would be used during BE tests (i.e., 200 m^3 h^{-1}). During that time, a portion of the biological sample collection system was used. One half of the sample collection system was used to determine whether the plankton nets for sampling organisms ≥ 50 µm would clog during uptake samples. In this test, no clogging occurred, and given that the source water had relatively high sediment and turbidity, the nets were not expected to clog during BE testing. It was noted that not all SOP procedures were completed during the sampling exercise. While sampling for organisms ≥ 10 µm and < 50 µm, a collection line burst during a backflush period. These minor incidents were addressed immediately, which minimized the potential for mishaps during a BE test. An additional problem was found: data from the BWMS were recorded with an incorrect timestamp because of factory firmware changes.

4.2.4 Summary from Commissioning

The installation and checkout at GBF was successful but required substantial modifications to the BWMS. These adjustments were accomplished and validated either by, or under the direction of, GBF engineering staff. Shakedown testing demonstrated the BWMS operated near its rated capacity of approximately 250 m^3 h^{-1} while flowing water. The average flow was less than this rate because the system had zero flow during backflush cycles. Tests with water with increased turbidity (visible by eye) showed additional reductions in average flow due to an increase in backflush frequency. It is noteworthy that both the control and treatment tank flow rates and the associated sampling operations appeared to be affected by the backflush operation of the BWMS. Furthermore, the resulting water hammer was deemed a safety concern and potential long-term reliability issue. The fact that a complete operational test was not conducted precluded the ability to perform a complete trial of all SOPs under a full operational scenario. The manufacturer was alerted to the water hammer concerns and potential remedies were discussed; this situation is further discussed in the next section. The manufacturer's technical representative also provided training to GBF staff. On 16 DEC 2010, GBF and the manufacturer's representative approved the commissioning of the BWMS. The successful commissioning resulted in updates to the facility SOPs and TQAP.

4.3 Golden Bear Facility Biological Efficacy Testing

A BE test in the GBF TQAP was defined as a full-scale ballast water test using water taken up from the Carquinez Strait at the vessel's mooring point in Vallejo, CA. All testing was performed using ambient water in brackish conditions, with the salinity and water quality parameters (e.g., the concentration of dissolved and particulate constituents) dependent on the mixing of the fresh water outflow from the Sacramento River and tidal inflow from San Pablo Bay. No augmentation of ambient water was performed. This ambient challenge water was pumped through the TF piping system from the sea chest and simultaneously split at a flow rate of ~200 m^3 h^{-1} to the control and treatment tanks. Challenge water was treated using the BWMS prior to entering the treatment tank but not the control tank.

4.3.1 Biological Efficacy Test 1 (10 – 14 JAN 2011)

Several issues were encountered during the first BE test. Prior to beginning testing, on 10 JAN 2011, GBF discovered during initial water flow through the BWMS that a damaged 6" (15.2 cm) socket weld flange had broken at the glue joint of the BWMS's PVC piping for the effluent water leaving the BWMS. The

joint had failed as the Lead Operator was shutting down the BWMS after performing a final checkout in preparation for the first BE test. Upon inspection, it was evident the flange failed at the glued joint to the 6" outlet butterfly valve downstream of the UV chamber. At the factory, the flange had been glued and successfully hydrostatically pressure tested, but under post-failure inspection, it appeared the joint had insufficient PVC adhesive to withstand the dynamic forces encountered when the BWMS went into backflush operation. A contractor hired by GBF repaired the broken flange, and the repair held for the remainder of the project.

Prior to the start of each test, the GBF team met to discuss testing activities and responsibilities. The ballast uptake operation was scheduled to start at 05:30 on Tuesday, 11 JAN 2011, to coincide with the beginning of ebb tide. Before starting the test, GBF personnel twice measured the water salinity, which was 9.3 and 9.6 psu. When GBF asked if they should proceed with the test because the readings were below the target value of 10 psu listed in the GBF TQAP, the VO agreed the test could proceed as the values were expected to rise over the course of the test.

The BE test started at 06:48, with water flow simultaneously directed to the control and treatment tanks. Sampling for organisms ≥ 50 µm started sequentially at 06:50 into all three sample tubs (each with its own plankton net; see Section *6.2.4.1 "Sampling Methodology — Golden Bear Facility"*). Sampling for organisms ≥ 10 µm and < 50 µm started at 06:55, after the carboy to collect the sample was rinsed. All of the sample water collected during uptake operations was obtained from sample port "S1", which was located in the piping upstream of the BWMS.

Uptake water was treated by the BWMS before being directed to the treatment tank. The BWMS was set to "Ballast" mode throughout the entire uptake cycle, and in this manner, water was directed first through the filtration system then through the UV reactor. According to the manufacturer's specifications, the filtration system was designed to filter all particles and organisms ≥ 50 µm nominally. As the filters loaded with sediment and debris, the BWMS initiated an automatic backflush procedure to reverse water flow to flush the filters.

As the test proceeded, the BWMS backflushed several times between the start of the test (06:48) and 07:30. At 07:30, the differential pressure (DP) sensor failed and read 0.0 psi for the remainder of the test. Because there was no indication of DP across the filtration system, GBF used the back-up procedure (as detailed in the TQAP) and initiated manual backflushes every 20 minutes. A violent (and loud) water hammer occurred when the "Outlet" butterfly valve closed during backflush operations, and there was concern another joint of the PVC piping might fail. In response to this issue, at 08:15, GBF personnel modified the operational procedure for the manual backflush operation: a GBF operator was assigned to manually open and close the 8" (20.3 cm) manual butterfly valve, which was located on the outlet side of the container. This change decreased the intensity of the shock but did not eliminate the water hammer.

Sampling for organisms ≥ 50 µm ended at 09:04 with a total volume of 3.6 m^3 (Sample Tank 1 volume: 1.21 m^3 [320 gallons], Sample Tank 2 volume: 1.20 m^3 [318 gallons], Sample Tank 3 volume: 1.20 m^3 [316 gallons]) as observed on the IMAC flow totalizer screens. After sampling ended, the individual plankton nets were rinsed from the exterior using water from the sample tubs; the two members of the science team were responsible for the procedure. After the rinse procedure was complete, the plankton net's cod end was disconnected, and the concentrated sample was poured into a glass beaker with 25 mL volumetric demarcations and ambient water (filtered with a 0.7 µm glass fiber filter [GF/F]) added to bring the volume

to 200 mL. The three samples appeared to be similar; all were opaque with black sediment. The sample processing procedure followed the GBF SOPs.

As samples for organisms ≥ 50 µm were collected, samples for the smallest two size classes were simultaneously collected from a tee in the same supply pipe. They were time-integrated, with sample water from the tee directed into three individual 20 L carboys. The samples were inverted several times prior to processing to ensure they were well mixed.

Immediately after collection, the biological samples were taken to the biology laboratory on the ship. The samples for organisms ≥ 50 µm were placed in a refrigerator until processing occurred (GBF SOPs did not specify temperature-controlled storage of samples; GBF was asked to store them in an incubator set to the challenge water's temperature to minimize thermal shock). In the initial observation of organisms ≥ 50 µm, GBF Analysts estimated approximately 40-50% of the organisms were living, whereas it was typical to find 70-80% of the organisms in this size class living in ambient water at the site. The majority of the organisms had to be gently prodded to determine if they were living (i.e., they responded to the mechanical stimulus). Copepod nauplii and veligers comprised the majority of the living organisms; tintinnids represented the majority of dead organisms. The samples were diluted 20-fold prior to analysis, and GBF Analysts said each 5 mL Bogorov chamber took approximately 10-12 min to evaluate.

The VO and AO witnessed GBF Analysts measuring the concentration of *Escherichia coli* using an enzyme-based most probable number (MPN) kit (IDEXX Colilert™); there was good agreement between Analysts on the color and degree of fluorescence in each well in the kit.

The drain operation of the treatment tank started at 11:06 on Thursday, 13 JAN 2011. As part of this operation, water was pumped from the treatment tank through the BWMS in "Deballast" mode (i.e., treated only by UV light) then overboard. After the treatment tank was completely drained, the control tank was drained in a similar manner, except water was not directed through the BWMS.

Sampling of the treatment tank water occurred as planned, and the TA and VO witnessed the GBF Science Team Lead decrease the sample flow rate to the sampling system as the main ballast flow dropped from ~200 $m^3 h^{-1}$ to ~50 $m^3 h^{-1}$ as "stripping" of the treatment tank occurred (i.e., removing the last ~12" [~30 cm] of water standing in the tank). The drain operation of the ballast tank was completed at 13:15.

Analysis of the biological samples from the treatment tank began immediately after the sample collection finished. As the results were tallied, GBF indicated the concentration of living organisms ≥ 50 µm in the first 5 mL Bogorov chamber was 36 living organisms, which was extrapolated to a density of 450 living organisms per m^3. Due to the higher-than-anticipated densities, treated samples were analyzed in the same manner as control samples (i.e., having a high density of organisms with a statistically normal distribution).

4.3.2 Biological Efficacy Test 2 (17 – 21 JAN 2011)

The BWMS vendor was contacted to address the issues of the water hammer and failed DP sensor from BE Test 1. After the call, the vendor supplied GBF with a new set of air-actuated bleed valves to be installed prior to the second test, as they would slow the valve actuation speed and thus reduce the water hammer. GBF installed the new valves but found this did not alleviate the problem, so GBF re-installed the previously used needle valves on the air supply.

Ballast uptake commenced on schedule on Tuesday, 18 JAN 2011 at 11:30. The BWMS was placed in "Ballast" mode, in which backflush was initiated by the DP across the filters. Because the DP readings were high (the GBF staff remarked that the sediment load appeared high), backflushes occurred frequently, as often as every 2 min. As the test progressed, the needle valves on the inlet and outlet valves were manually adjusted by GBF Test Operations Staff to slow the opening and closing speed of the valve to minimize the water hammer. After several backflush cycles, the closing speed of the output valve was successfully adjusted, although the output valve seemed to stick or pause during opening and did not operate smoothly.

Early in the treatment tank uptake operation, GBF Test Operations Staff observed high DP readings following each backflush; it appeared to require a minute or two for the readings to return to a lower value. In an attempt to prevent the temporarily high DP reading from triggering another backflush, GBF personnel tried removing the DP sensor cable (which should have resulted in a zero reading and thus no backflush). However, after twice blowing fuses as the cable was reconnected (which resulted in a shutdown of the BWMS), this procedure was deemed unacceptable. GBF Test Operations Staff then initiated manually timed backflushes. These started at 10-min intervals for two cycles and then increased to 15-min intervals, while checking to ensure that minimum desired flow was maintained. Because the DP sensor on the BWMS was inoperable, the facility used flow rate and time as the metric to determine when to initiate a manual backflush. If the flow rate through the BWMS started dropping below the TQAP minimum of 200 m^3 hr^{-1}, it indicated the BWMS filter had sufficient particle load to require a manual backflush. In addition, the BWMS was designed with the option to operate in one of two modes: 1) initiation of backflush based on DP across the filter, or 2) initiation of backflush based on a pre-determined time interval. The use of a manual backflush was considered consistent with system operation, and the timing was selected to prevent occurrence of high DP across the filter. This procedure allowed the test to be completed without further interruptions and without dropping the total uptake flow rate below 360 m^3 h^{-1}. Afterwards, the volumes in the control and treatment tanks were approximately 400 m^3 in the treatment tank and 395 m^3 in the control tank, within the ranges specified in the GBF TQAP.

Because of the problems with the valve actuators and the DP sensor, the BWMS manufacturer was contacted to provide on-site support for this test. A technical representative arrived on Wednesday morning, 19 JAN 2011, and remained on site for the remainder of the week and into the third week of BE testing to inspect repairs and the operational status of the BWMS. It was the VO's opinion that the issues listed above did not affect treatment capabilities as long as the backflush was manually controlled (which was the case throughout all of the BE tests). The BWMS's backflush problems appeared to be primarily due to a faulty DP transducer component.

The discharge operations occurred on Thursday, 20 JAN 2011. Flushing of the sample collection equipment commenced at 09:05, and the treatment discharge started at 09:35, finishing at 11:35. The control discharge operation started at approximately 13:35 and continued to 15:35. Treatment and control discharges were relatively smooth, and the small discrepancies observed were minor delays in the startup or adjustment of sampling. For instance, during the treatment discharge, GBF Science Team was not aware that discharge had started, so there was a 1-2 minute delay before sampling commenced. Also, the initial flow setting for the sampling manifold overflow (dump) was set to the 3 x 1 m^3 setting (8.4 gpm); this was corrected to the 10 m^3 setting (2.5 gpm) within the first five min. The GBF staff noticed these errors relatively quickly and, from an operational standpoint, they were viewed by GBF and the VO (who was on-site during the test) as inconsequential.

During the end of the treatment tank's deballasting (with flow reduced to 50 m^3 h^{-1} to strip the tank), the BWMS was observed going in and out of warning mode several times. Furthermore, UV intensity dropped during this time, at one point as low as 50% (the target minimum is 70%). This drop coincided with a drop in transmissivity of the water to 1%. Clearly, stripping the tank did, indeed, capture settled sediment within the tank and re-suspended it into the flow (a qualitative comparison showed the final water discharged from the tank to have much more particulate material than the initial discharge). Because GBF collected biological samples over the entire discharge period, (as per the ETV Protocol) it would be impossible to determine if this event had any effect on the results during this specific time.

4.3.3 Biological Efficacy Test 3 (24 – 28 JAN 2011)

At the final BE test, the TA was present. Two phone conferences between the TA, the VO, and GBF were conducted, after the uptake and discharge operations. During the uptake operation (which began at 05:29 on Tuesday, 25 JAN 2011 and ended at 07:42), GBF indicated they replaced two failed DP sensors and had to revert to performing manual backflushes for the second half of the test. Engineering parameters were met during the uptake cycle, but the science team had difficulty completing all of the analysis and preparations in time for the discharge operation. The Analysts worked until 04:00 the next day, and the *Vibrio cholerae* analyses were completed the following day.

The discharge operation (of the treatment tank) started at 09:19 on Thursday, 27 JAN 2011 and ended at 11:20. The control discharge began at 12:48 and was stopped at 14:50. GBF indicated they observed a small 0.5 cm tear in the mesh of one of the treatment tank plankton nets, which could allow organisms to escape, uncounted. The GBF Science Team thought they would have noticed a tear on the other two tests because of its location on the net, therefore, they were confident that the tear did not impact results of the first two BE tests. The GBF SOPs did not include an inspection of the plankton nets prior to each use.

4.4 Golden Bear Facility Results

4.4.1 Golden Bear Facility Executive Summary

The ETV Protocol lists 'Core Parameters', which are the minimum measurements required needed to verify the validity of a BE Test Cycle:

- Temperature
- Salinity
- Total suspended solids
- Particulate organic matter
- Dissolved organic matter
- Dissolved oxygen
- pH
- Ambient organism concentration
- Ballast system flow rate
- Ballast system pressure
- Sampling flow rates
- Chlorophyll *a* (biomass)

All core parameters were measured by GBF during the BE tests. One item of mention is the ballast system pressure (line pressure) on the inlet for the BWMS. While ballast system pressure was measured, the sensor was located in the engine room (several decks below the BWMS), so the measurements did not reflect the pressure at the inlet (and thus could not be compared to the manufacturer's specifications). While the ETV Protocol does not specify the location of the sensor, it is desirable to locate the sensor at the BWMS inlet, and GBF has added this capability for future BWMS testing. Organisms in the ≥ 10 µm and < 50 µm size class were analyzed with two methods, epifluorescence microscopy and flow cytometry.

The summarized results from the three BE tests conducted at GBF are shown in Table 3. Most of the entries in the table were copied directly from the GBF Executive Summary (Golden Bear Facility, 2012). Additional parameters were added to the table to summarize the control uptake and discharge ballast line pressures, treatment uptake and discharge line pressures, and the dissolved oxygen (DO) concentration measured during all tests. These data were compiled from GBF Appendix C-Data Logs Automation and from Table 6-11 of the verification report (Golden Bear Facility, 2012). The data in this table are discussed below in Section 6 "Comparison of Protocol Execution and Results between Test Facilities".

Table 3. Golden Bear Facility data summary.

Parameters and units (method, where applicable)		TQAP Criteria	BE 1[A] 11 JAN 2011	BE 2 18 JAN 2011	BE 3 25 JAN 2011	Subsamples per cycle
Uptake Conditions						
Uptake Water Chemistry						
Salinity (psu)		10 - 20	**8.3** - 19.3	13.6 - **20.1**	11.8 - 18.7	Ongoing through cycle
Temperature (°C)		4 - 35	8.4 - 9.9	10.0 - 12.4	10.3 - 11.0	Ongoing through cycle
Dissolved organic carbon, DOC (mg L^{-1})	Mean (SD)	≥ 3	**2.1** (0.06)	**2.1** (0)	**2.3** (0.06)	3
Particulate organic carbon, POC (mg L^{-1})	Mean (SD)	≥ 2.0	**1.129** (0.0742)	**1.187** (0.0175)	**1.155** (0.0339)	3
Total suspended solids, TSS (mg L^{-1})	Mean (SD)	≥ 20	47.3 (2.39)	67.6 (1.31)	65.2 (3.27)	9
Chlorophyll *a* (µg L^{-1})	Mean (SD)	NR	0.84 (0.05)	1.27 (0.04)	1.35 (0.10)	9
pH		NR	7.87	7.85	7.80	1
Dissolved oxygen (mg L^{-1})		NR	10.0 - 11.1	9.7 - 10.3	9.8 - 10.1	Ongoing through cycle
Uptake Living Organisms						
Organisms ≥ 50 µm (organisms m^{-3})	Mean (SD)	≥ 10,000 > 5 species/ > 3 phyla	86,300 (11,000) > 5 species/ > 3 phyla	18,900 (2150) > 5 species/ > 3 phyla	37,100 (4050) > 5 species/ > 3 phyla	18
Organisms ≥ 10 and <50 µm (organisms mL^{-1}) (epifluorescence microscopy)	Mean (SD)	≥ 1000 > 5 species/ > 3 phyla	**320** (65.5) > 5 species/ > 3 phyla	1990 (284) > 5 species/ > 3 phyla	**757** (97.0) > 5 species/ > 3 phyla	9
Organisms ≥ 10 and <50 µm (organisms mL^{-1}) (flow cytometry)	Mean (SD)	≥ 1000 > 5 species/ > 3 phyla	**340** (53)	**690** (92)	**660** (99)	9
Escherichia coli (cfu 100 mL^{-1})	Mean (SD)	NR	7 (3.4)	20 (7.3)	12 (3.5)	9
Enterococci (cfu 100 mL^{-1})	Mean (SD)	NR	78 (21)	150 (43)	129 (43)	9

Table 3. Golden Bear Facility data summary (Continued).

Parameters and units (method, where applicable)		TQAP Criteria	BE 1[A] 11 JAN 2011	BE 2 18 JAN 2011	BE 3 25 JAN 2011	Subsamples per cycle
Uptake Conditions						
Uptake Living Organisms - Continued						
Organisms < 10 µm (culturable, aerobic, heterotrophic bacteria using HPC) (cfu mL^{-1})	Mean (SD)	≥ 1000	1700 (1300)	2350 (1390)	**994** (466)	9
Vibrio cholerae (O1/O139) (cfu 100 mL^{-1})		NR	< 1	< 1	< 1	3
Uptake Ballast Operations						
Treatment tank volume – End of cycle (m^3)		400 ± 10%	432.0	401.8	403.7	NA
Treatment line flow – Average (backflush time excluded) (m^3 h^{-1})		200 ± 10%	201.7	202.1	202.7	NA
Control tank volume – End of cycle (m^3)		400 ± 10%	362.6	394.5	392.9	NA
Combined sample volume for organisms ≥ 50 µm (m^3)		≥ 3	3.6	3.7	3.6	3
Control line flow – Treatment average (backflush time included) (m^3 h^{-1})		NR	162.3	170.9	181.5	NA
Control inlet manifold pressure (kPa)	Mean (SD)	NR	299.0 (94.0)	317.7 (85.5)	344.3 (67.5)	NA
Combined sample volume for organisms ≥ 10 µm and < 50 µm (L)		NR	60	60	60	3
Control Discharge Conditions						
Living Organisms						
Organisms ≥ 50 µm (organisms m^{-3})	Mean (SD)	100	55,200 (7920)	30,100 (3530)	50,000 (4610)	18
Organisms ≥ 10 µm and < 50 µm (organisms mL^{-1}) (epifluorescence)	Mean (SD)	100	703 (185)	958 (295)	413 (87.1)	9
Organisms ≥ 10 µm and < 50 µm (organisms mL^{-1}) (flow cytometry)	Mean (SD)	100	191 (36.2)	397 (142)	344 (65.0)	9
E. coli (cfu 100 mL^{-1})	Mean (SD)	NR	8 (5)	2 (1)	2 (1)	9
Enterococci (cfu 100 mL^{-1})	Mean (SD)	NR	9 (5)	103 (43)	75 (9)	9
Organisms < 10 µm (culturable, aerobic, heterotrophic bacteria using HPC) (cfu mL^{-1})	Mean (SD)	500	**409** (356)	**259** (141)	511 (196)	8
Vibrio Cholerae (O1/O139) (cfu 100 mL^{-1})		NR	< 1	< 1	< 1	3
Water Chemistry						
Chlorophyll *a* (µg L^{-1})	Mean (SD)	NR	0.37 (0.04)	0.76 (0.11)	0.61 (0.05)	9
pH		NR	7.83	7.85	7.84	1
Ballast Operations						
Control tank volume – End of cycle (m^3)		NR	18.4	12.7	11.9	NA
Control line flow – Average (m^3 h^{-1})		200 ± 10%	202.1	201.4	204.6	NA
Control line discharge pressure (kPa)	Mean (SD)	NR	113.8 (51.4)	120.7 (52)	52 (47.4)	NA
Combined sample volume for organisms ≥ 50 µm – Control discharge (m^3)		≥ 3	**2.9**	3.3	3.2	3
Combined sample volume for organisms ≥ 10 µm and <50 µm – Control discharge (L)		NR	60	60	60	3

Table 3. Golden Bear Facility data summary (Continued).

Parameters and units (method, where applicable)		TQAP Criteria	BE 1[A] 11 JAN 2011	BE 2 18 JAN 2011	BE 3 25 JAN 2011	Subsamples per cycle
Treatment Discharge Conditions						
Living Organisms						
Organisms ≥ 50 µm (organisms m^{-3})	Mean (SD)	NR	127 (31)	217 (62)	284 (72)	18
Organisms ≥ 10 µm and < 50 µm (organisms mL^{-1}) (epifluorescence)	Mean (SD)	NR	9.1 (1.8)	6.5 (0.55)	4.4 (1.5)	4
Organisms ≥ 10 µm and < 50 µm (organisms mL^{-1}) (flow cytometry)	Mean (SD)	NR	21 (11)	12 (14)	17 (14)	9
E. coli (cfu 100 mL^{-1})		NR	< 1	< 1	< 1	9
Enterococci (MPN 100 mL^{-1})		NR	< 1	< 1	< 1	9
Organisms < 10 µm (culturable, aerobic, heterotrophic bacteria using HPC) (cfu mL^{-1})	Mean (SD)	NR	4 (5)	150 (120)	8 (10)	9
Vibrio Cholerae (O1/O139) (cfu 100 mL^{-1})		NR	< 1	< 1	< 1	3
Water Chemistry						
Chlorophyll *a* (µg L^{-1})	Mean (SD)	NR	0.39 (0.05)	0.41 (0.16)	0.48 (0.06)	9
pH		NR	7.75	7.83	7.82	1
Ballast Operations						
Treatment tank volume – End of cycle (m^3)		NR	7.7	7.5	7.6	NA
Treatment line flow – Average (m^3 h^{-1})		200 ± 10%	202.9	203.4	199.0	NA
Treatment line discharge pressure (kPa)	Mean (SD)	NR	158.8 (35.5)	142 (32.3)	118.2 (50.9)	NA
Combined sample volume – Treatment Discharge (m^3)		≥ 9	9.6	9.3	9.5	3
Combined sample volume for organisms ≥10 µm and <50 µm – Treatment discharge (L)		60	60	60	60	3
Ballast hold duration (h)		48 ± 10%	52	46	52	NA

[A]Values shown in bold font and in shaded cells indicate the minimum concentration defined in the GBF TQAP was not met. BE = biological efficacy, cfu = colony forming unit, HPC = heterotrophic plate counts, NA = not analyzed, NR = no requirement, psu = practical salinity units, SD = 1 standard deviation, and TQAP = test quality assurance plan.

4.4.2 Golden Bear Facility Deviation Matrix

In any process as complicated as BE testing, it is not surprising that deviations from the TQAP occurred. The deviations that arose during testing and GBF identified in the verification report are listed in Table 4. They are accompanied by the actions GBF took to address the problem in subsequent tests, GBF's interpretation of the deviation's potential impact on testing, and comments by the VO. Deviations are discussed in depth in the GBF verification report, Section 7.6, "Weekly Discussion of Test Cycles" (Golden Bear Facility, 2012); the other deviations identified in the table were noted subsequent to the two BE tests witnessed by the VO as discussed below.

Table 4. Deviations from the Golden Bear Facility Test Quality Assurance Plan.

BE test(s)	Deviation	Corrective action taken by GBF to address the issue in the future	Outcome on test results per GBF	Comments by VO
1	During treatment of water upon uptake, following the BWMS backflush cycles the outlet valve on the BWMS opened too quickly causing a loud water hammer and visible pipe flex; to safely continue the test, GBF engineers slowly, manually opened the outlet valve	Installed speed controls, which were provided by the vendor, on the valve actuators; this action largely eliminated the water hammer	Not addressed specifically by GBF in the verification report	VO approved of the action on-site and determined the effect on test results was likely negligible
1-3	During treatment of water upon uptake, the DP sensor on the BWMS failed, and in order to continue the test, GBF engineers manually initiated backflushes in lieu of the BWMS DP sensor automatically initiating backflushes	Installed new DP sensors, which were provided by the vendor, for the next BE tests (i.e., tests 2 and 3)	No effect, as backflushing would have occurred at least as often if the DP sensor had been working properly	VO approved of the action on-site or in advance and concurred that no effect on test results was likely
1-3	Following the hold time and recirculation of treated water prior to discharge, flash rust developed in the steel piping; to minimize the rust in samples and staining of plankton nets, the discharge line was flushed for 60 s before sampling commenced	No action available	Minimal, as the flushing accounted for ~3 m^3, approximately 1% of the treatment discharge	VO approved of the action on-site and agreed the effect on test results was likely minimal
1	Upon discharge, some treated water was recirculated through the suction pump because two valves were not tightly closed; halfway through the discharge cycle, the valves were completely closed	None	No effect, as double-block and bleed isolation at each tank prevented leakage	VO agreed no contamination likely occurred
2	During treatment of water upon uptake, a fuse was blown causing the BWMS to shut down for a total of 30 min; a likely cause was a power lug that grounded to the plug mounting screw	The screw was removed on the DP sensor so it did not cause an electrical short	No effect, as water was recirculated within the pipe loop during this short time	VO agreed there was likely no effect on test results
1-3	Grab samples were not analyzed for salinity, temperature, conductivity, and dissolved oxygen with a handheld probe per SOP 11, and there was no comparison of the grab samples to thermosalinograph data per SOP 23; this deviation was not discussed in the verification report	Not discussed in the verification report; later discussions with GBF showed that in lieu of comparing grab samples to the thermosalinograph, the data collected using the thermosalinograph was compared to a data sonde, which had an unknown calibration history, deployed approximately 30 m from and 2 m lower than the GBF sea chest intake	Not discussed	Because the sonde to which the thermosalinograph was compared had an unknown calibration history, the data are a qualitative assessment of water quality

Table 4. Deviations from the Golden Bear Facility Test Quality Assurance Plan (Continued).

BE test(s)	Deviation	Corrective action taken by GBF to address the issue in the future	Outcome on test results per GBF	Comments by VO
1-3	SOPs in the verification report describing DOC (SOP 1.4) and TSS (SOP 1.2) were not followed (see Table 17 for details)	Not discussed	Not discussed	In post-test discussions with the VO, the SOPs used in testing were described, although samples were not stored according to ETV Protocol guidelines; thus the data are questionable
1-3	SOP 26, the chain of custody form, was not used; this deviation was not discussed in the verification report	Not discussed	Not discussed	Records from the contracted off-site lab and email correspondence were provided by GBF, but this procedure needs to be used for future testing

BE = biological efficacy test, BWMS = ballast water management system, DP = differential pressure, ETV = Environmental Technology Verification program, GBF = Golden Bear Facility, VO = Verification Organization, and SOP = standard operating procedure.

The most significant deviations were the backflushing and water hammer issues, which were ameliorated by the GBF engineering team, who worked quickly so testing could continue on schedule. Criteria used to indicate that manual initiation of the backflushing cycles was appropriate are discussed in the GBF verification report, Section 7.6.1, "Test Cycle 1 Uptake and Validation of Test Parameters" (Golden Bear Facility, 2012). The vendor provided assistance by sending new DP sensors and speed controls for the valve actuators on the BWMS, although the former worked intermittently throughout the subsequent two tests. It appeared that none of these issues affected the biological or water quality data collected. The manual initiation of cyclical backflushing did not confound the results of the test with regard to the project objective (i.e., ability to run ETV tests). Because the DP sensor was not functional, the BWMS could not be operated in automatic mode; rather it was operated with a timed, manual backflush, as approved by the manufacturer. Furthermore, a technical representative from the BWMS manufacturer was on-site during commissioning (where the water hammer also occurred) and during the second and third BE tests, and he deemed the BWMS to be in working order.

The GBF verification report provided a few SOP summaries described as "updated to reflect small changes that might have been incorporated during the execution of the three test experiments", but there was no discussion of specific changes or their significance. Upon review of the redlined SOPs, the VO observed other deviations that were not discussed in the GBF verification report. A review of the TQAP and associated SOPs in Appendix A (dated 30 APR 2012) to the GBF verification report showed that redlines were made months after completion of the tests (as late as 01 JUN 2011). Two deviations from the TQAP are included in Table 4; they were not conducted during testing and subsequently marked as "not used". These were the analysis of grab samples using the handheld measurement probe (and subsequent comparison to inline thermosalinograph data), and the chain-of-custody forms for transporting samples to an off-site laboratory. Other minor changes to SOPs were also noted and are discussed elsewhere in this report.

5 GREAT SHIPS INITIATIVE

5.1 Great Ships Initiative Test Quality Assurance Plan

The following sections describe the process of developing the TQAP and associated documentation at GSI.

5.1.1 Required Elements of the Test Quality Assurance Plan

The ETV Protocol lists requirements that should be included in the TQAP; they are listed above in Section 4.1.1 "Golden Bear Facility Test Quality Assurance Plan". The GSI TQAP was assessed by the VO and AO as part of the review and acceptance process, and it was compliant with all elements.

5.1.2 Development Process

The QAPP was provided by GSI prior to their writing of the TQAP at the VO's request. Both the VO and AO reviewed the preliminary GSI QAPP in January 2011, discussed it on a conference call with GSI on 28 JAN 2011, and provided written feedback to GSI on 04 FEB 2011. Recommendations were provided on how to incorporate the QAPP into the TQAP and SOPs, along with general and specific suggested changes for adherence to the ETV Protocol requirements. The VO also discussed with GSI that results and data should not be shared between the Intercomparison Project TFs, even though they did communicate at the time about other issues. A complete draft set of test documentation was provided by GSI on 22 FEB 2011.

Afterwards, a conference call was held with GSI to discuss the preparation of QAPP and TQAP documents and the schedule for preparing documents and testing (24 FEB 2011).

Both the VO and AO reviewed GSI's draft test documentation, which they discussed with GSI during a conference call on 09 MAR 2011. Written feedback on the TQAP, QMP, and QAPP (combined comments from both the AO and VO) was provided to GSI on 14 MAR 2011, and written comments on the SOPs from NSF followed on 17 MAR 2011. GSI responded to the comments with a set of revisions, questions, and commentary on the written feedback through documents provided on 15 APR 2011. This submission was followed by a conference call between the VO and AO on 21 APR 2011 to discuss the TQAP, QMP, and QAPP; additional written feedback on these items was provided to GSI on 25 APR 2011. Questions from GSI on the SOPs were received on 27 APR 2011; these were discussed in a conference call on 29 APR 2011 between the VO and AO, and followed by written feedback to GSI on the same day.

On 23 MAY 2011, GSI provided a final draft set of documents. A series of minor comments was provided as feedback on 01 JUN 2011, with a request for a final update to occur subsequent to the commissioning of the BWMS in early June. This process allowed any revisions resulting from the commissioning activities to be incorporated into the TQAP and SOPs. All updated documents were provided by GSI (subsequent to commissioning) on 14 JUN 2011.

Developing GSI's TQAP (and associated documents) required and benefitted from the allocation of time to allow several revisions and both the written and verbal interactions between the VO, AO, and the GSI teams. Although two major drafts were prepared prior to the final version, the last revision was largely to add clarifications and incorporate any changes resulting from commissioning. Given that the TF had never before developed this level of documentation, the time and effort expended seemed reasonable.

5.2 Great Ships Initiative Commissioning

The BWMS was successfully installed and commissioned at the GSI facility in Superior, WI during the first two weeks of June 2011. Relatively minor problems were encountered during commissioning, and all were discussed with GSI staff and rectified by the end of the week. Prior to testing, a checklist was prepared and the relevant commissioning SOP reviewed; this was similar in content to that used with GBF. All relevant tasks were observed on-site by representatives from the VO and TA. During this trip, the VO and TA also witnessed GSI personnel perform a full-scale, BE test (as part of the commissioning) to ensure organism and sediment injection operations functioned properly, sampling operations yielded the correct water volumes, and the BWMS operated properly at the facility. As part of the practice test, GSI staff evaluated the concentrations of living organism and water chemistry parameters to ensure they met ETV Protocol requirements. Finally, as part of the commissioning, several operational changes were identified for GSI test procedures; primarily these involved equipment setup or setting of operational parameters. These were subsequently incorporated by GSI in the final revision of the TQAP. The commissioning was witnessed by two representatives from the VO and one TA representative from USCG-RDC. It was also attended by technical representatives from the BWMS vendor (due to travel schedules, attendance was split between two people). The technical representatives provided training to GSI staff and approved of the commissioning.

5.2.1 Installation

The BWMS container was installed on the floor within Laboratory Building 1. GSI personnel installed the 6" (15.2 cm) ballast inlet, 6" (15.2 cm) outlet, 4" (10.2 cm) backflush, air supply, and power connections. The TF used schedule 80 PVC pipes to connect their piping system to the BWMS system. Visual inspection

confirmed the piping, electrical system, and compressed air supply were installed properly, all valves operated and moved per the control signal sent to them, the discharge booster pump rotated properly, and the communication cables were installed properly.

GSI chose not to interface control or alarm signals from the BWMS system to the facility's control and data logging system, with the exception of one interlock signal. GSI preferred to use the supplied remote panel interface supplied with the BWMS and manually monitor this for status as well as provide commands to the system. In part, this was due to limited free space available on their graphical display and touch screen as well as a preference to limit custom programming on their control system.

5.2.2 Checkout and Operational Tests

Prior to any operations, GSI completed the SOP form for safety procedures during testing. The GSI Facility Operations Manager gave instructions to all parties involved in or witnessing the commissioning regarding safety, exits, and the emergency eye wash stations. He also indicated that anyone on the site could use the emergency stop switch located in the GSI control room if he or she saw impending danger.

Initial pressure tests were performed with water flows at 300 gpm and pressure up to 35 psi to check for leaks. These were performed prior to delivering power to the BWMS system. After powering up the system, the manufacturer's technical representative determined that both a power transformer and line voltage settings were incorrectly configured. The manufacturer corrected these problems, but a small 24 VDC power supply failed, likely because the supply voltage switch on the power supply was initially set to 110VAC instead of 220VAC. This was successfully replaced. The manufacturer also determined that the UV transmittance sensor was incorrectly calibrated. The sensor was recalibrated (twice) on site. It should be noted that this sensor was not checked during commissioning at GBF, as it was factory-calibrated, and the operations manual did not include instructions for calibrating the sensor.

After the sensor was changed, the vendor agreed to add steps to the manual for calibrating the sensor upon start up. However, the BWMS control system appeared to consistently skip the intermediate of three possible power levels when adjusting power to the UV lamps, instead of sequencing through the power levels as designed. Due to the high levels of DOC in ambient water at GSI, it was anticipated that the system would require full power from the lamps for the tests. Regardless, the manufacturer's technical representative applied a software change to the BWMS prior to the first BE Test at GSI, ensuring the UV level sequenced properly from power level 1 to 3 without skipping power level 2.

During initial backflush testing, two small piping leaks were discovered at the flange where the BWMS mated with the GSI facility piping system. Both leaks were rectified by facility personnel by tightening the bolts on the leaking flanges. Subsequently, GSI performed a full-scale uptake and discharge of both the treatment and control tanks at a flow rate of approximately 200 $m^3 h^{-1}$ to each tank. The BWMS was operated in 'Ballast' mode on uptake, and was allowed to automatically control backflushes based on readings from the DP sensor. GSI used their Sediment Injection System (SIS) and Organism Injection System (OIS) to inject water chemistry constituents (harbor water augmented with test dust and humic material) and biological organisms (collected in plankton tows the previous day). The goal of the test was to execute sampling and injection operations and monitor flow between the treatment and control tracks leading to the treatment and control tanks while meeting challenge water conditions outlined in the ETV Protocol. This generated backflush conditions typical of the BE tests (as discussed below). The VO and TA observed this process to compare GSI's activities with the TQAP and facility SOPs. These activities included test operations, sampling activities, biological analyses, and pre- and post-test grooming activities.

5.2.3 Summary from Commissioning

The BWMS was successfully commissioned at GSI on 09 JUN 2011, with the installation approved by both GSI and the BWMS vendor. The VO and TA witnessed the commissioning process as described by GSI in their facility SOPs. There were minor issues with the BWMS that, although unresolved, were deemed by the VO and GSI as unlikely to affect the Intercomparison test activities. These appeared to result from work performed at the factory where the BWMS was sent after the GBF test (see Section 6.3.1 "Commissioning"), and also included the sequencing of the UV power control (skipping the intermediate level). As discussed above, a software change was applied to the BWMS prior to the first BE Test at GSI to ensure the UV level sequenced properly from power level 1 to 3 without skipping power level 2. The major issues associated with the previous test, severe water hammer from inadequate valve damping and the failure of the DP sensor, appeared to be resolved. Although still present at GSI, the water hammer was minimal (minimal noise, no deflection of piping, and no shaking of the container during backflush operations), and the DP sensor operated normally with no failures during commissioning. GSI was able to successfully install and operate the system and ran an operational test under typical challenge conditions. Their commissioning test did not include operating the BWMS at its maximum capacity of 250 $m^3 h^{-1}$ flow (not required in the ETV Protocol), but it did include full BE uptake and discharge operations per the GSI TQAP. This practice run not only resulted in updates to the facility SOPs and TQAP, but it also demonstrated the facility's ability to successfully operate and test the BWMS as planned.

5.3 Great Ships Initiative Biological Efficacy Testing

A BE test was defined at this facility as a full-scale ballast water test using water taken up from the Duluth-Superior Harbor of Lake Superior. All testing was performed using ambient freshwater augmented with (1) organisms collected from the harbor (intended to augment organisms ≥ 10 µm and < 50 µm, as described below) and injected from the facility's OIS and (2) POC and MM injected from the facility's SIS. Water was pumped from the facility intake pump through a single intake line, augmented with organisms and material from the OIS and SIS, and then was simultaneously split to the control and treatment tracks, each with a flow rate of approximately 200 $m^3 h^{-1}$. Challenge water was treated by the BWMS prior to entering the treatment tank but not the control tank. Water was held in the control and treatment tanks for a period of approximately 48 h, and the treated water was discharged through the BWMS to the harbor, followed by the control water, which was discharged directly to the harbor.

5.3.1 Biological Efficacy Test 1 (12 – 14 JUL 2011)

Prior to testing, GSI described the process used to augment natural concentrations of organisms ≥ 10 µm and < 50 µm: University of Wisconsin-Superior (UWS) students (under the guidance of the Zooplankton Team Lead) collected organisms by towing a 0.5 m diameter plankton net with 50 µm mesh from a small boat in Duluth-Superior harbor. The net was towed just below the surface for 10 - 15 minutes or until the net clogged such that it no longer filtered the water. This process collected protists, which were often filaments or other colonial forms. The effective pore size of the net was quickly reduced as it became progressively clogged, such that tows also captured organisms ≥ 50 µm. The plankton tows, which took several hours, occurred one or two days prior to an uptake operation. Following collection, the organisms were stored in one of two small holding ponds situated outside the GSI TF that were fitted with aerators and mixers. Temperature and oxygen levels were monitored in the tank so the physiological state of the organisms could be estimated the morning of an uptake operation. The water (and organisms) in the holding ponds were then injected into the ballast line during an uptake operation through the OIS, which used a diaphragm pump to provide flow and pressure.

On the day of an uptake operation, GSI performed biological analysis prior to the start of the operation, to determine if adequate concentrations of organisms ≥ 50 µm were present in the ambient water to meet the requirements of the test to be performed. Because GSI did not have the capability to augment organisms in this size class, if the concentration was low, GSI waited 2 h and assessed the density of organisms again. If, after the two-hour wait period, organism densities were still low, the test would be postponed until the following day. GSI indicated that if this situation were to occur during the Intercomparison Project, GSI would discuss the course of action with the VO.

During each week of BE testing, on Monday mornings, GSI distributed to the testing team a schedule of tasks and personnel to conduct them. Some tasks were as short as five minutes, so the testing was finely choreographed. The planned uptake time for the first BE test was 10:00.

At 09:35, after completing an initial analysis of a sample of the harbor water, the Zooplankton Team Lead indicated there were sufficient organisms ≥ 50 µm in the ambient water to start the test. The uptake operation commenced when the Facility Operations Manager started the intake pump at 10:06. After the piping system pressurized, one of the protist sample hoses came loose from its designated carboy. Within about thirty seconds, a Staff Engineer secured the hose and reinstalled it in the carboy. With the BWMS set to 'Ballast' mode, the first automatic backflush of the filter occurred at 10:18, and GSI observed the DP reading across the filter decrease from 11 psi to 2.1 psi after the backflush was completed, indicating the filter cleaned itself successfully. The water hammer issue observed at GBF was much less evident here, and only a small amount of flexing in the piping within the BWMS container was noted.

At 10:37, GSI indicated the BWMS was backflushing more frequently than they initially had planned, and with the VO and TA concurrence, the flow rate to the control tank was reduced so the treatment and control tanks would fill evenly. At 11:05, the VO was informed that an oversight by GSI engineers had occurred: the flow rate to the control tank had been entered into the facility control system incorrectly. The correct flow rate was reentered into the control system, and the discrepancy was noted in the operator test data sheet.

At 11:09, the hose connecting the OIS came loose during a backflush operation of the BWMS. The hose loosened at the union between the injection pipe and main ballast line. The Operations Manager was informed and quickly shut the ball valve on the diaphragm pump and the ball valve on the sample port. Within approximately one minute, he reconnected the hose and reopened the valves. This event was noted in the operator test data sheet.

The fill operation was completed at 11:26. Next, the VO observed the processing of samples for organisms in the two smallest size classes (≥ 10 µm and < 50 µm; < 10 µm). The samples for organisms ≥ 10 µm and < 50 µm were collected in 20 L carboys, which were inverted by a GSI biologist 5 times. The samples in the carboys were then poured into 1 L plastic sample bottles. The sample bottles for organisms ≥ 10 µm and < 50 µm were rinsed with water from the carboys. The sample bottles for organisms < 10 µm, which had been sterilized, were not rinsed, as indicated by the SOP for this operation. Both sets of samples were placed in a cooler, taken to the appropriate laboratory, and processed immediately following the GSI SOP.

The VO then watched the processing of the samples for organisms ≥ 10 µm and < 50 µm in the laboratory at GSI. Initially, the Phytoplankton Team Lead attempted to filter 200 mL of sample through a 7 µm sieve but found the samples were too dense and needed diluting. After dilution, he followed the GSI SOP for sample analysis. The pre-printed data sheets included photomicrographs of the different groups of organisms

commonly encountered in GSI samples. GSI Analysts used microscopes to count and record the living number of cells and used an ocular micrometer to measure the cells. When diatoms were seen, the Phytoplankton Analyst recorded the number of cells in a chain on the datasheet; chain-forming diatoms dominated the samples analyzed that day. The number of living cells < 10 µm was also recorded. GSI indicated they had a time window of 1.5 h for processing the samples. One of the GSI Quality Assurance and Quality Control (QA/QC) Officers was present during the analysis.

Next, the VO witnessed the processing of samples for the largest size class (organisms ≥ 50 µm). The treatment uptake samples were processed first, followed by the control uptake samples. The Sampling Staff member in charge of this operation followed the GSI SOPs with no deviations. Processing of an uptake treatment sample began at 11:30. The 35 µm plankton net was suspended above a 130 L plastic barrel. A hose connected to the drain spout of the sample collection tub (which was elevated above the plankton net) was routed through the mouth of the plankton net to the cod end at the bottom of the net. In this manner, the water flowing from the sample tub was dispensed underwater, into the plankton net, and thus the water (and organisms entrained in it) did not fall freely through the air to hit either the bottom of the barrel or the surface of the water. Next, the entire contents of the sample-collection tub were drained through the plankton net, emptying the tub by 11:48. Afterwards, the Sampling Staff member rinsed the inside of the sample collection tub using a hose that was connected to a submersible pump at the bottom of an adjoining plastic barrel, thus rinsing the tub with water that had been filtered through the plankton net (i.e., filtrate). After that, he moved back to the plankton net and followed the SOP for pulling the net out of the plastic barrel. He rinsed all four quadrants of the net from the exterior, working from top to bottom, using filtrate water, and then removed the net's cod end. The cod end was capped, placed in a cooler, and taken to the zooplankton laboratory for processing. The Zooplankton Team Lead examined the sample to determine if dilution was necessary.

On Wednesday, 13 JUL 2011, the VO was present when GSI cleaned the facility piping system using the "Hotsy™" (http://www.hotsy.com/index.aspx) system, which used pressure washing and potable water from the city of Superior heated to 160 °F (71.1 °C). GSI cleaned the piping system in sections. First, sampling ports on the piping system were opened, the Hotsy™ hose nozzle was inserted through an open port, and the hose was fed into the piping to slowly move the nozzle forward through the section of pipe until it reached the end of the manifold or a stopping point. Next, GSI used potable water from the city of Superior to flush the entire portions of the piping system that were shared during uptake and discharge operations at a rate of 1,000 gpm for 15 min. As the potable water circulated, GSI opened all sample ports to flush the sample tubs for organisms ≥ 50 µm and all hoses used for discharge sampling. On 14 JUL 2011, prior to the treatment discharge, a 1 m^3 sample of potable water was collected in a randomly selected sampling tub for treatment samples and then filtered through a plankton net as described above to determine if there were any remaining organisms ≥ 50 µm in the water. GSI also examined the potable water at its source, as they previously had seen contamination from freshwater invertebrates: nematodes (worms), rotifers (genus *Lecane*), and chironomids (midge larvae). If any of these organisms were found while testing the piping, they were considered "exclusions", because they had been found occasionally in the TF's potable water system. All water was drained from the piping system from its lowest point. On this day, one living organism ≥ 50 µm (not one of the "exclusions") was discovered after this cleaning procedure. The flushing process was then repeated two times. (If any of these "exclusions" organisms were present in treated water collected in this study, their presence would be irrelevant, given the number of other organisms in this size class).

On Thursday, 14 JUL 2011, starting at 09:19, GSI sequentially drained the treatment tank and then the control tank. The BWMS was set to 'Deballast' mode prior to start of this operation, and the flow rate through the BWMS was approximately 200 $m^3 h^{-1}$. The VO watched the GSI Staff Engineer and Engineering Intern remove data sondes, one from the control tank, one from the treatment tank, which had been deployed during the hold time to measure temperature, conductivity, salinity, dissolved oxygen, pH, turbidity, and total chlorophyll. Sampling proceeded as planned, and the main treatment ballast flow was stopped at 10:16.

The VO observed a 64% UV intensity level on the BWMS during much of this operation, indicating the water was highly turbid. According to the BWMS manual, when the UV intensity drops below 70%, the display screen for the UV system will provide a warning that reads "Low Intensity"; because this screen is not the default screen, the user must select "UV System Information and Control" from among the 11 available screens to view this information. The "Low Intensity" condition will initiate successive increases in lamp power, up to maximum power at level 3. The BWMS manual indicates that "Low Intensity" warns the "UV Sensor is sensing less than acceptable levels of UV intensity" and requires the following actions: "Check that the lamp breaker(s) have not tripped. Check life period of lamp. Check condition of ballast water for fouling agent, i.e., oil, shallow ballast draw, etc." There is no instruction to halt operations or that the treatment may be insufficient. During testing for this project, the lamps had low operating hours, were operating normally, and the ballast intake location was at a fixed height well clear of the harbor bottom; turbidity was due to the ambient water conditions. As anticipated, the system ran at level 3 during testing. As no alarm signal, system shutdown, or indicator requiring operator response was provided under these conditions, testing continued under the assumption that the BWMS operated normally and within design specifications.

Sample processing for the ≥ 50 µm size class started at 10:25. The nets were suspended above the 130-L plastic barrels, and sampling proceeded as described in GSI's SOP. One minor problem occurred with the water used to rinse the nets after the cod cup (with the sample) had been removed from the net (another cod cup had been put on the net). GSI collected this water on the treatment discharge to count the number of organisms that may have remained attached to the inside of the net until water filtered through the plankton net was used to rinse the nets from the outside. When a Sampling Staff member was processing a sample of this rinse water, he accidentally overfilled the 1-L sample bottle and lost approximately 5% of the water. This sample was not the main treatment sample; rather it was from the net rinse water, therefore neither GSI nor the VO was concerned. The SOP was revised to ensure that any future overflowed water would be retained in a 4.7 L pail and collected in an additional sample bottle.

Next, the GSI Zooplankton Team explained their sample processing procedure to the VO. GSI indicated they combined the contents of the cod end and the water collected from rinsing the plankton net (described in the paragraph above). The sample was poured into a Folsom plankton splitter. Half of the sample was analyzed using a compound microscope for the smaller organisms in the ≥ 50 µm size class commonly found at GSI, generally rotifers, copepod nauplii, and mussel veligers (collectively called "microzooplankton"). The other half of the sample was analyzed using a dissecting microscope for the larger organisms in the ≥ 50 µm size class commonly found at GSI, generally adult copepods, cladocerans, and other macroinvertebrates (collectively called "macrozooplankton"). The samples were then each poured into an ungraduated glass sample jar and concentrated (if the population was rare) to reduce the volume of the sample. Next, the Zooplankton Team Lead performed an initial check of the microzooplankton sample by loading 1 mL of sample into a Sedgewick Rafter (SR) counting chamber. In this instance, fifty swimming (i.e., living) organisms were counted in one row of the SR, which was

extrapolated to a density of approximately 47,000 organisms m^{-3}. The several species of rotifers present were dominated by the genus *Keratella*. The Analysts examining the macrozooplankton samples found one living oligocheate worm. From this initial estimation of the treatment discharge sample, the Zooplankton Team Lead indicated GSI would perform the ≥ 50 µm analysis as if the samples were control uptake or control discharge samples, that is, having a high density of live organisms with a statistically normal distribution. The control tank discharge commenced at 13:22 and continued until it was stopped at 14:20. The sample processing also proceeded as planned.

As analysis for organisms ≥ 50 µm was conducted at the GSI site, the VO went to the nearby UWS campus to view the processing of the < 10 µm samples by the two GSI Microbial Analysts. All sample plates and vessels had pre-printed labels, and the three incubators in use were set at 42 °C (107.6 °F), 56 °C (132.8 °F), and 36 °C (96.8 °F), with each having remote temperature probes to collect daily records of the minimum and maximum temperatures. The sample processing was consistent with the SOPs.

5.3.2 Biological Efficacy Test 2 (27 – 29 JUL 2011)

GSI scheduled the second test to begin the week of 18 JUL 2011, but unusually high temperatures resulted in very low organism concentrations in the ambient water at that time. As a result, the second test was not started until organism concentrations had increased sufficiently.

On Wednesday, 27 JUL 2011, the concentrations of both of the largest size classes (≥ 50 µm; ≥ 10 µm and < 50 µm) were met or were within 10% of the GSI target values, meeting the requirements in the GSI TQAP. The start of the test was delayed until 11:23 to ensure that appropriately high concentrations of organisms were in the uptake water. This test lasted 104 minutes, ending at 13:07, which was considerably longer than the first test. The test ran longer because the BWMS backflushed about every three minutes, which meant the flow rate to the treatment tank was considerably slower than the first test.

Because of the longer fill time, the OIS tanks for organisms ≥ 10 µm and < 50 µm collected from the harbor were emptied approximately 80% of the way through the uptake cycle, thus additional organisms were not injected into the challenge water for the final 20% of the test period. GSI documented this issue, and the VO was not concerned because the facility collected time-averaged samples for biological and water quality analyses. After the OIS tanks of collected organisms were emptied, the backflush frequency dropped by a factor of approximately 2.5 - 3 based on the observed data for flow rate and pressure.

On Thursday, 28 JUL 2011, the GSI staff cleaned and flushed the piping system in preparation for Friday's discharge operation. The VO observed part of the cleaning operation.

GSI conducted a successful discharge operation on Friday, 29 JUL 2011. The discharge operation started first with the treatment tank and then with the control tank; all operations proceeded as planned.

5.3.3 Biological Efficacy Test 3 (09 – 11 AUG 2011)

The GSI uptake and discharge operations for BE Test 3 were observed by one member of the VO staff. The uptake operation started on Tuesday, 09 AUG 2011 at 10:03, and the BWMS backflushed approximately every five minutes. In response, the Operations Manager adjusted the main ballast flow rate to the control tank, sample tubs, and injection flow rates. The uptake cycle was stopped when GSI discovered the valve that drains to the harbor from the main ballast uptake line had been left open. Instead of sending all of the water to the treatment and control tanks, some of the water had been sent to the harbor. This oversight was discovered at approximately 10:41. Because of this error, the test was aborted and a new test was initiated

at 11:58 using the two other treatment and control tanks at the facility. The uptake cycle was completed at 13:35. The sample processing and analysis operations proceeded as documented in GSI's SOPs.

GSI delayed the discharge operation slightly on Thursday, 11 AUG 2011 because a plastic cap on the bottom of an experimental fluorescence instrument fell off as the unit was being lowered into the control tank in an exercise unrelated to the BE test. GSI was concurrently testing this system as a side project with the Science Team Lead at GBF (note that both GSI and GBF had confidentiality agreements in place and agreed not to discuss their respective tests). GSI decided to search for and remove the cap from the tank because it could damage the impeller on the discharge ballast pump. To extract the cap, GSI located it with an underwater camera and used a fishhook and line to pull it out of the tank. After this delay, the discharge operation proceeded as planned. The operation started at 10:26 with the treatment tank and stopped at 11:22. GSI followed the facility SOPs for sample processing and analysis, and the control tank discharge operation was conducted with no issues started at 14:20 ending at approximately 15:30.

5.4 Great Ships Initiative Results Summary

5.4.1 Great Ships Initiative Executive Summary

The ETV Protocol lists the 'Core Parameters', which are the minimum requirements needed to verify the validity of a BE Test Cycle (for a list of the parameters, see Section 4.4.1 "Golden Bear Facility Executive Summary"). Nearly all core parameters were measured by GSI during the BE tests. The results of the three BE tests conducted at GSI are shown in Table 5, which was populated with information contained in the Executive Summary of the GSI verification report (Great Ships Initiative, 2012). Total chlorophyll data were not reported in the Executive Summary of the GSI verification report, but because they were reported in Table 22 of the GSI report, the values were added to Table 5. Water chemistry measurements for control and treatment tank discharges for DOM, DOC, MM, and TSS were not obtained during the first BE test, as these measurements were not included in the original TQAP developed by GSI. This oversight was pointed out to GSI by the VO after the first BE Test, and GSI added these measurements for the final two tests. Finally, GSI did not report the values for *E. coli* or Enterococci in the Executive Summary of the verification report, but the data were found in the GSI raw data files, so the mean values for the control uptake, treatment uptake, control discharge, and treatment discharge are reported in Table 5.

Table 5. Great Ships Initiative data summary.

Parameters and units (method, where applicable)		TQAP Criteria (method)	BE 1[A] 12 JUL 11	BE 2 25 JUL 11	BE 3 09 AUG 11
Uptake Conditions					
Uptake Water Chemistry					
Salinity (ppt)		≤ 1	0.09	0.09	0.09
Temperature (°C)		4-35	22.40	20.82	23.48
Dissolved organic matter (mg L^{-1} as DOC)	Mean (SD)	≥ 6	18.7 (0.8)	12.3 (0.4)	16.3 (0.1)
Particulate organic matter (mg L^{-1} as POC)	Mean (SD)	≥ 4	5.9 (0.8)	5.0 (1.3)	4.0 (0.6)
Mineral matter (mg L^{-1})	Mean (SD)	≥ 20	24.6 (3.2)	21.4 (1.0)	**17.7 (1.5)**
Total suspended solids (mg L^{-1})	Mean (SD)	≥ 24	30.4 (2.4)	26.5 (0.7)	**21.7 (2.0)**
pH		6-9	7.4 (0.1)	7.4 (0.1)	7.5 (0.1)
Total chlorophyll (µg L^{-1})		NR Control NR Treatment	14.5 14.4	9.8 9.3	15.6 14.9
Dissolved oxygen (mg L^{-1})		6-11	7.44	7.57	6.35

Table 5. Great Ships Initiative data summary (Continued).

Parameters and units (method, where applicable)		TQAP Criteria (method)	BE 1[A] 12 JUL 11	BE 2 25 JUL 11	BE 3 09 AUG 11
Uptake Conditions					
Uptake Living Organisms					
Organisms ≥ 50 µm[B] (organisms m^{-3})	Mean	≥ 10,000 > 5 species/ > 3 phyla	341,000 > 6 species/ 3 phyla	353,000 > 6 species/ 3 phyla	396,000 > 6 species/ 3 phyla
Organisms ≥ 10 and < 50 µm[C] (epifluorescence microcopy) (organisms mL^{-1})	Mean	NA	605 > 5 species/ > 3 phyla	572 > 5 species/ > 3 phyla	560 > 5 species/ > 3 phyla
Organisms ≥ 10 and < 50 µm[D] (epifluorescence microcopy) (organisms mL^{-1})	Mean	≥ 1000 > 5 species/ > 3 phyla	1530 > 5 species/ > 3 phyla	1490 > 5 species/ > 3 phyla	2340 > 5 species/ > 3 phyla
Escherichia coli (cfu 100 mL^{-1})	Control (Mean)	NR	158	382	997
	Pre-Treatment (Mean)		157	559	543
Enterococci (cfu 100 mL^{-1})	Control (Mean)	NR	28	180	1047
	Pre-Treatment (Mean)	NR	55	123	380
Organisms < 10 µm (culturable, aerobic, heterotrophic bacteria, mL^{-1})	Mean (SD)	≥ 1000 MPN (IDEXX SimPlates®)	10,000 (4000)	190,000 (30,000)	86,700 (30,600)
	Mean (SD)	≥ 1000 cfu (spread plates)	54,100 (12,900)	140,000 (48,000)	137,000 (79,000)
Uptake Ballast Operations					
Treatment tank volume – End of Cycle (m^3)		175-205	198	193	192
Treatment line flow – Average, with backflush time excluded (m^3 h^{-1})[E]		180-210 (excludes backflushes)	197	**177**	184
Pre-Treatment Tub 4 sample flow rate (m^3 h^{-1})		2.0-3.6	3.0	2.7	2.9
Control Tub 1 sample flow (m^3 h^{-1})		2.0-3.6	2.3	**1.8**	2.1
Control tank volume – End of cycle (m^3)		175-205	193	185	192
Combined sample volume for organisms ≥ 50 µm in control and treatment tanks (m^3)		2.6-3.6	2.6 (control) 3.0 (treatment)	3.1 (control) 2.9 (treatment)	3.2 (control) 3.0 (treatment)
Control line flow (m^3 h^{-1})		140-170	164	**108**	**126**
Pressure – Control and treatment track split (bar and psi)		2.4-3.4 (35-50 psi)	2.7 (39 psi)	2.9 (42 psi)	3.0 (43 psi)
Combined sample volume for organisms ≥ 10 µm and < µm – (L)		19	15 (control) 15 (treatment)	19 (control) 19 (treatment)	15 (control) 16 (treatment)
Retention time in control and treatment tanks (h)		43.2-52.8	49.9 (control) 45.9 (treatment)	47.7 (control) 43.6 (treatment)	48.9 (control) 44.9 (treatment)
Control Discharge Conditions					
Living Organisms					
Organisms ≥ 50 µm[B] (organisms m^{-3})	Mean	≥ 100	413,000	593,000	559,000
Organisms ≥ 10 and < 50 µm[C] (epifluorescence microscopy) (organisms mL^{-1})	Mean	≥ 100	448	389	674
E. coli (cfu 100 mL^{-1})	Mean	NR	16	62	199
Enterococci (cfu 100 mL^{-1})	Mean	NR	500	57	2016
Organisms < 10 µm (culturable, aerobic, heterotrophic bacteria, mL^{-1})	Mean (SD)	≥ 500 MPN (IDEXX SimPlates®)	7300 (460)	40,000 (20,000)	113,000 (16,000)
	Mean (SD)	≥ 500 cfu (spread plates)	32,300 (4600)	99,600 (15600)	79,400 (8900)

Table 5. Great Ships Initiative data summary (Continued).

Parameters and units (method, where applicable)		TQAP Criteria (method)	BE 1[A] 12 JUL 11	BE 2 25 JUL 11	BE 3 09 AUG 11
Control Discharge Conditions					
Water Chemistry					
Salinity (ppt)		NR	0.09	0.09	0.09
Temperature (°C)			20.58	22.05	23.18
Dissolved organic matter (mg L^{-1} as DOC)	Mean (SD)		NM[F]	11.9 (0.05)	17.2 (0.4)
Particulate organic matter (mg L^{-1} as POC)	Mean (SD)		NM	2.2 (1.9)	2.0 (1.0)
Mineral matter (mg L^{-1})	Mean (SD)		NM	12.3 (1.4)	12.6 (1.1)
Total suspended solids (mg L^{-1})	Mean (SD)		NM	14.6 (2.1)	14.6 (0.2)
pH			7.69	8.18	7.22
Total chlorophyll (µg L^{-1})			12.0	8.5	13.0
Dissolved oxygen (mg L^{-1})			6.67	6.98	5.35
Ballast Operations					
Control tank volume – End of cycle (m^3)		NR	NR	NR	NR
Control line flow – Average (m^3 h^{-1})		190-210	199	199	200
Control line discharge pressure (bar, with psi in parentheses)		2.1-2.8 (30-40 psi)	2.1 (30 psi)	2.1 (30 psi)	2.1 (31 psi)
Combined sample volume for organisms ≥ 50 µm – Control discharge (m^3)		3.3-3.6	**3.7**	3.6	**3.7**
Combined sample volume for organisms ≥ 10 µm and < 50 µm – Control discharge (L)		NR	NR	NR	NR
Treatment Discharge Conditions					
Living Organisms					
Organisms ≥ 50 µm[G] (organisms m^{-3})	Mean (SD)	NR	77,600 (13,400)	25,400 (5810)	48,100 (3,890)
Organisms ≥ 10 and < 50 µm[C] (epifluorescence) (organisms mL^{-1})	Mean	NR	301	197	95
E. coli (cfu 100 mL^{-1})	Mean	NR	5	8	43
Enterococci (cfu 100 mL^{-1})	Mean	NR	6	23	1711
Organisms < 10 µm (culturable, aerobic, heterotrophic bacteria, mL^{-1})	Mean (SD)	NR MPN (IDEXX SimPlates®)	10,400 (2200)	15,800 (3400)	19,900 (8700)
	Mean (SD)	NR cfu (Spread Plates)	36,900 (8700)	33,900 (8400)	33,700 (1900)
Water Chemistry					
Salinity (ppt)	Mean (SD)	NR Mean (SD)	0.09 (0.00)	0.09 (0.00)	0.09 (0.00)
Temperature (°C)	Mean (SD)		20.89 (0.03)	21.58 (0.05)	22.94 (0.04)
Dissolved organic matter (mg L^{-1} as DOC)			NM[F]	12.2 (0.3)	17.0 (0.5)
Particulate organic matter (mg L^{-1} as POC)			NM	1.9 (1.1)	1.3 (2.2)
Mineral Matter (mg L^{-1})			NM	10.1 (0.8)	13.0 (4.0)
Total suspended solids (mg L^{-1})			NM	12.0 (0.4)	14.3 (2.5)
pH			7.63 (0.07)	7.78 (0.02)	7.40 (0.03)
Total chlorophyll (µg L^{-1})			11.7 (0.2)	6.9 (0.1)	11.9 (0.2)
Dissolved oxygen (mg L^{-1})			6.94 (0.01)	7.28 (0.02)	5.79 (0.02)
Ballast Operations					
Treatment tank volume (m^3) – End of cycle		NR	NR	NR	NR
Treatment line flow (m^3 h^{-1}) – Average		190-210	195	197	197
Treatment line discharge pressure (bar and psi)		2.1-2.8 (30-40 psi)	2.2 (32 psi)	2.2 (32 psi)	2.2 (32 psi)
Combined sample volume – Treatment discharge (m^3)		3.3-3.6	**3.7**	3.4	3.6
Combined sample volume for organisms ≥ 10 µm and < 50 µm – Treatment discharge (L)		57	46.5	42.0	46.5
Ballast hold duration (h)		43.2-52.8	45.9	43.6	44.9

[A]Values shown in bold font and in shaded cells indicate the minimum specification in the GSI TQAP was not met. [B]Organisms collected for this size class in the uptake and control discharge samples were *typically* ≥ 50 µm in minimum dimension (Great Ships Initiative, 2012, TQAP Appendix 6, SOP GSI/SOP/LB/RA/SA/2) and may have included some organisms < 50 µm in minimum dimension. [C]Organisms were measured according to the ETV Protocol of ≥ 10 µm and < 50 µm in "maximum dimension on the smallest axis" (EPA, 2010); here, values without SDs represent a single, composited sample. [D]Organisms were measured according to the DFS (≥ 10 µm and < 50 µm on any axis). [E]The target mean flow rates differed between the control and treatment tracks due to backflushing of the BWMS in the treatment track, which caused a pause in the flow of water. The target mean flow rate in the control track was slowed so the control and treatment tanks were filled at the same time and had the same final volume. [F]As per the GSI TQAP, water quality grab samples were not collected during control or treatment discharge. After Test Cycle 1, the TQAP was revised to include the collection of these samples. [G]At GSI, all living organisms in treatment discharge samples are usually measured, but in this study, because the density of organisms was higher than typically found in treated samples, a subsample of organisms was measured, and the proportion of organisms < 50 µm in the sample was extrapolated so the final, reported densities excluded organisms < 50 µm. BE = biological efficacy, cfu = colony forming unit, DOC = dissolved organic carbon, MPN = most probable number, NA = not applicable, NM = not measured, NR = no requirement, POC = particulate organic carbon, psi = pressure per square inch, ppt = parts per thousand, SD = 1 standard deviation, and TQAP = test quality assurance plan.

At GSI, the community of zooplankton is typically dominated by rotifer populations with size distributions that can straddle the minimum size requirement of ≥ 50 µm size class and the maximum size requirement of the ≥ 10 µm and < 50 µm size class. In the uptake and control discharge water, which had high concentrations of organisms, measuring all organisms was impracticable. In validation studies, GSI has found the number of rotifers < 50 µm, typically from the genera *Keratella* and *Polyarthra*, does not affect the TF's ability to meet challenge water conditions. In previous tests of BWMSs, the treatment discharge samples contained low densities of organisms, so GSI scientists took a photomicrograph of each living organism and determined its size from the image. In this project, the relatively high density of living organisms in treated water precluded that approach, so in BE tests 2 and 3, a subsample of living organisms was photographed (while living) and the maximum dimension on the smallest axis was determined. From a total of 105 usable images of *Keratella* and 17 usable images of *Polyarthra*, the number of organisms < 50 µm was estimated (9.5% of the *Keratella* and 29.5% of the *Polyarthra*; Great Ships Initiative, 2012). These percentages were applied to the samples collected from each BE test, thereby adjusting the reported densities of organisms ≥ 50 µm to exclude organisms < 50 µm. The data in Table 5 are further discussed in Section 6 "Comparison of Protocol Execution and Results between Test Facilities".

5.4.2 Great Ships Initiative Deviation Matrix

Deviations from the TQAP occurred at GSI, and they are listed in Table 6 along with the actions GSI took to address each problem in subsequent tests as well as GSI's interpretation of the deviation's potential impact on testing and comments by the VO (for a complete discussion of the deviations, see the GSI verification report, Section 8.1.2 "Intake and Discharge Operations/Engineering" through Section 8.1.7 "Organisms ≥ 50 µm" [Great Ships Initiative, 2012]). No other deviations were noted during the two BE tests that were witnessed by the VO, and those that GSI identified (as per Table 6) were minor.

Table 6. Deviations from the Great Ships Initiative test quality assurance plan.

BE Test(s)	Deviation	Corrective action taken by GSI to address the issue in the future	Outcome on test results per GSI	Comments by VO
1	During the uptake for the control tank, the sample flow rate was too high and thus not proportional to the main line flow; when the error was discovered, the flow rate was changed through a series of adjustments	The SOP was adjusted to insure the issue did not re-occur	Minimal, as the uptake samples for the control tank samples may not have been representative of the uptake water, but the treatment flow rate was on target	VO concurred that the effect on test results was minimal
1-3	In-line, continuous pH measurements were not collected on the uptake; to collect pH data, a total of 3 grab samples were taken from the sample port where the sensor would have been, with one sample collected at the beginning, middle, and end of the uptake operation	3 grab samples were collected for all tests, and pH was measured in the pre-treatment sample collection tub after uptake	Minimal, as grab samples were collected	VO concurred that the effect on test results was minimal
1	The temperature of samples was not recorded, nor was the pH 8.00 Check Buffer analyzed after all samples were measured, on the uptake or discharge	The deviation was detected in the post-test audit, discussed with the chemist, and corrected on next tests	Minimal, as the pH meter was calibrated prior to analysis each day, and it is unlikely to have fallen out of calibration in two days	VO concurred that the effect on test results was minimal
1	Equipment used to prepare samples for *Vibrio cholerae* colony blots was not sprayed with ethanol between replicate samples from the control uptake, pretreatment uptake, or treatment discharge; to determine the extent of the cross contamination, between the last 2 treatment discharge replicates, sterile, deionized water was filtered through a new filter, and the second treatment discharge replicate was reanalyzed after equipment was cleaned with ethanol	Used ethanol to disinfect equipment in subsequent tests	Minimal, as there was no difference between the results from the second treatment discharge replicate when ethanol was used and when it was not	VO concurred that the effect on test results was minimal, and all results were zero
1	In verifying the efficacy of tank cleaning prior to the BE test, the number of dead zooplankton in several categories were not recorded on data sheets	The GSI QA/QC Officer (or another GSI research team member) will double-check the datasheets for completion	None; the SOP requires the count of living organisms and the tally of dead organisms if time permits; no living organisms were found	VO concurred that there was no effect on result
1-3	During analysis of treatment discharge samples, chironomid larvae (insects commonly known as midges) were counted, and they were excluded from zooplankton counts (as they are sometimes found in the potable water used to clean the pipes at GSI)	The SOP should have incorporated this practice, as a validation study in 2011 by GSI showed chironomids were a natural source of contamination	None	VO concurred that this practice should be performed and incorporated into the SOP

Table 6. Deviations from the Great Ships Initiative test quality assurance plan (Continued).

BE Test(s)	Deviation	Corrective action taken by GSI to address the issue in the future	Outcome on test results per GSI	Comments by VO
2	No water quality parameters were measured continuously and *in situ* in the control and treatment tanks during the 48-h hold time; after discharge of the tanks, a data sonde was deployed in the sample tubs, and the data indicated all water quality parameters were within acceptable ranges at the time of discharge	Corrected in the next test	Minimal, as the data sonde indicated the water quality was acceptable	VO concurred that the effect on test results was minimal
2	Control discharge samples for *Enterococcus* spp. and *E. coli* were placed in the incorrect incubators; when the deviation was noticed, all control and treatment discharge samples, which had been stored in a refrigerator, were re-analyzed and placed in the proper incubator	The correct incubator was used in the next test	Results may have been lower than expected because of the holding time in the refrigerator	VO concurred that this event may have reduced counts
NA	In the proposed BE Test 2, GSI staff were not able to collect enough phytoplankton for the injection ponds, and ambient concentrations were too low to meet challenge conditions	The test was postponed until phytoplankton densities were sufficient	The testing window was lengthened beyond the original plan	The testing window was lengthened but was still within the agreed upon limit
2	Some of the phytoplankton in the injection ponds were > 2 d old when the test commenced	Densities were sufficient in the next test	None, as the number of phytoplankton were sufficient for the test because new cells compensated for older, dying cells (dead cells were not counted)	Although some older cells may have been less robust than cells <2 d old, the numbers in the control tank discharge met ETV requirements; the effect was undetectable but likely minimal

BE = biological efficacy, GSI = Great Ships Initiative, NA = not applicable, VO = verification organization, and QA/QC = quality assurance and quality control.

6 COMPARISON OF PROTOCOL EXECUTION AND RESULTS BETWEEN TEST FACILITIES

6.1 Challenge Conditions

6.1.1 Water Quality Characteristics

The ETV Protocol specifies water quality conditions at TFs (Table 7) so that BWMSs are tested under challenging conditions and also to simulate potential water quality characteristics that may be encountered in real-world, shipboard operations. As discussed earlier, the two TFs participating in this project, GBF and GSI, are located on different water types, brackish water and freshwater, respectively. A TF situated at a marine site was not tested as part of this project. Furthermore, DFSs were given to both TFs because they did not meet some of the ETV Protocol requirements (Table 2).

The results for water quality measurements collected during both TFs' BE tests are shown in Table 7. Although the ETV Protocol states that water quality measurements should be obtained for both the uptake and discharge operations, GBF only reported water quality data on the uptake operation using a thermosalinograph and did not provide grab sample analyses of water quality parameters per the GBF TQAP. Uptake and discharge data were provided for TSS and chlorophyll *a*; however, POC and DOC measurements were not performed. Unfortunately, these deviations were not discovered by the VO and GBF until after all GBF testing was concluded. Interestingly, the uptake measurements for TSS were approximately two times higher than the values measured in the control tank discharge and approximately 25-50% higher than values in the treatment tank discharge. These values seemed high, even considering settling that would occur and incomplete draining of the tank. While statistical analyses were not performed on these data, they seem to warrant further investigation as part of a TF validation. At GSI, the originally approved TQAP did not call for collection of water quality grab samples during control or treatment discharge. After Test Cycle 1, the GSI TQAP was revised to include the collection of these samples.

Once testing began, there were several instances when GBF was unable to meet the minimum uptake parameters for DOC and POC. Because the facility relied upon ambient water conditions and had no ability to augment it, the VO, with ESAC concurrence, accepted the test conditions to allow continuation of the project. Additionally, in two tests at GBF (Tests 1 and 2), the salinity fell outside the range specified in the TQAP. The VO allowed the tests to continue because GBF had no ability to change the water quality parameters, and the salinity was expected fall within the range over the majority of the test. In nearly all instances, GSI exceeded the minimum targets for uptake water as specified in the TQAP for DOM, POM, MM, and TSS. The only exceptions were for MM and TSS for Test 3 (Table 7).

Table 7. Water quality requirements from the Environmental Technology Verification Protocol and results from biological efficacy tests.

Protocol or TF Operation	Water quality parameters: Requirements and results from GBF and GSI BE Tests[A]					
	Water type (Salinity) Results[B]	**Dissolved organic matter** Results (SD)	**Particulate organic matter** Results (SD)	**Mineral matter** Results (SD)	**Total suspended solids** Results (SD)	**Temperature** Results
ETV Uptake req	Fresh (< 1 psu) Brackish (10 - 20 psu)	6 mg L^{-1} as DOC	4 mg L^{-1} as POC	20 mg L^{-1}	24 mg L^{-1}	4 – 35 °C
GBF Uptake req	Brackish (10 - 20 psu)	3 mg L^{-1}	2 mg L^{-1}	18 mg L^{-1}	20 mg L^{-1} 1	4 – 35 °C
BE 1	**8.3** - 19.3	**2.1** (0.06)	**1.129** (0.0742)	46.171	47.3 (2.39)	8.4 - 9.9
BE 2	13.6 - **20.1**	**2.1** (0.00)	**1.187** (0.0175)	66.413	67.6 (1.31)	10.0 - 12.4
BE 3	11.8 - 18.7	**2.3** (0.06)	**1.155** (0.0339)	64.045	65.2 (3.27)	10.3 - 11.0
GBF Dis req	NR	NR	NR	NR	NR	NR
BE 1	NM	NM	NM	NM	22.3 (5.7)	NM
BE 2	NM	NM	NM	NM	27.5 (1.1)	NM
BE 3	NM	NM	NM	NM	29.8 (2.8)	NM
GBF TT Dis req	NR	NR	NR	NR	NR	NR
BE 1	NM	NM	NM	NM	38.0 (2.4)	NM
BE 2	NM	NM	NM	NM	48.7 (8.5)	NM
BE 3	NM	NM	NM	NM	44.7 (3.3)	NM
GSI Uptake req	Fresh (< 1 ppt)	≥ 6 mg L^{-1}	≥ 4 mg L^{-1}	≥ 20 mg L^{-1}	≥ 24 mg L^{-1}	4 – 35 °C
BE 1	0.09	18.7 (0.8)	5.9 (0.8)	24.6 (3.2)	30.4 (2.4)	22.40
BE 2	0.09	12.3 (0.4)	5.0 (1.3)	21.4 (1.0)	26.5 (0.7)	20.82
BE 3	0.09	16.3 (0.1)	4.0 (0.6)	**17.7** (1.5)	**21.7** (2.0)	23.48
GSI CT Dis req	Fresh (< 1 ppt)	NR	NR	NR	NR	NR
BE 1	0.09	NM	NM	NM	NM	20.58
BE 2	0.09	11.9 (0.5)	2.2 (1.9)	12.3 (1.4)	14.6 (2.1)	22.05
BE 3	0.09	17.2 (0.4)	2.0 (1.0)	12.6 (1.1)	14.6 (0.2)	23.18
GSI TT Dis req	Fresh (< 1 ppt)	NR	NR	NR	NR	NR
BE 1	0.09	NM	NM	NM	NM	20.89
BE 2	0.09	12.2 (0.3)	1.9 (1.1)	10.1 (0.8)	12.0 (0.4)	21.58
BE 3	0.09	17.0 (0.5)	1.3 (2.2)	13.0 (4.0)	14.3 (2.5)	22.94

[A]Values shown in bold font and in shaded cells indicate the ETV or TQAP requirement was not met. [B]Standard deviations were not calculated for this measurement. BE = biological efficacy, CT Dis = control tank discharge, DOC = dissolved organic carbon, DOM = dissolved organic matter, ETV = Environmental Technology Verification program, GBF = Golden Bear Facility, GSI = Great Ships Initiative, MM = mineral matter, NM = not measured, NR = no requirement, POC = particulate organic carbon, POM = particulate organic matter, psu = practical salinity units, req = requirement from the ETV Protocol or the facility's TQAP, SD = 1 standard deviation, TQAP = test quality assurance plan, and TT Dis = treatment tank discharge.

Another way to examine the water quality conditions and gauge a facility's ability to test BWMS according to the ETV Protocol, *without any DFSs*, is to calculate the concentration a given constituent (e.g., POC) in the uptake water as a percentage of the ETV Protocol minimum requirement (Table 8).

Table 8. Test Facilities' uptake water chemistry results compared to the minimum requirements of the Environmental Technology Verification program Protocol.

Parameter	Minimum ETV Requirement	GBF Water Chemistry Concentrations[A, B]	GSI Water Chemistry Concentrations
Dissolved organic carbon ($mg\ L^{-1}$)	6	BE 1: **2.1 (35%)** BE 2: **2.1 (35%)** BE 3: **2.3 (38%)**	BE 1: 18.7 (312%) BE 2: 12.3 (205%) BE 3: 16.3 (272%)
Particulate organic carbon ($mg\ L^{-1}$)	4	BE 1: **1.1 (28%)** BE 2: **1.2 (30%)** BE 3: **1.2 (29%)**	BE 1: 5.9 (148%) BE 2: 5.0 (125%) BE 3: 4.0 (100%)
Mineral matter ($mg\ L^{-1}$)	20	BE 1: 46.2 (231%) BE 2: 66.4 (332%) BE 3: 64.0 (197%)	BE 1: 24.6 (123%) BE 2: 21.4 (107%) BE 3: **17.7 (89%)**
Total suspended solids ($mg\ L^{-1}$)	24	BE 1: 47.3 (197%) BE 2: 67.6 (282%) BE 3: 65.2 (272%)	BE 1: 30.4 (127%) BE 2: 26.5 (110%) BE 3: **21.7 (90%)**

[A]Values shown in bold font and in shaded cells indicate the ETV requirement was not met (i.e., the value is < 100%).
[B]GBF mineral matter concentrations were calculated from reported values of total suspended solids and particulate organic carbon. ETV = Environmental Technology Verification, GBF = Golden Bear Facility, and GSI = Great Ships Initiative.

Viewing the data in this manner shows that neither TF could meet all of the ETV water quality requirements in all three tests. GBF, which did not use augmentation, met approximately one third of the DOC and POC concentrations, and exceeded MM (and TSS, the sum of MM and POC) requirements by two- to three-fold. GSI, using augmentation for POC and MM (and, by extension, TSS), exceeded most of the requirements in most of the tests, with the DOC exceeded the most (two- to three-fold higher than the ETV challenge water requirement).

6.1.2 Concentrations of Living Organisms in Challenge Water

The ETV requirements, requirements for the TFs for this project, and data from BE Tests are listed in Table 9. Note that the pathogenic and indicator bacteria are excluded from the table, as there is no requirement for their concentrations in the ETV Protocol. Considering the challenge water conditions outlined in the ETV Protocol (without DFSs), none of the tests conducted under this project would have met all requirements for the three size classes. Potential changes to the challenge water requirements are discussed in Section 7 "Suggested Changes to the Environmental Technology Verification Protocol and Comparison to International Maritime Organization G8 Guidelines".

Table 9. Challenge water requirements for living organisms and results from biological efficacy tests.

Protocol or TF operation	Requirements for challenge water conditions and results from GBF and GSI[A]		
	Requirements **Living organisms** **≥ 50 µm** **Diversity requirement** **Results** **Mean (SD)** **Diversity**	**Requirements** **Living organisms** **≥ 10 µm and < 50 µm[B]** **Diversity requirement** **Results** **Mean (SD)** **Diversity**	**Requirements** **Living organisms** **< 10 µm[C]** **Diversity requirement** **Results** **Mean (SD)** **Diversity**
ETV	≥ 100,000 m^{-3} ≥ 5 species from ≥ 3 phyla	≥ 1,000 mL^{-1} ≥ 5 species from ≥ 3 phyla	≥ 1,000 mL^{-1} NR
GBF **Uptake** **BE 1** **BE 2** **BE 3**	≥ 10,000 m^{-3} ≥ 5 species from ≥ 3 phyla 86,300 (11,100) m^{-3} 18,900 (2,150) m^{-3} 37,100 (4,050) m^{-3} ≥ 5 species from ≥ 5 phyla	≥ 1,000 mL^{-1} ≥ 5 species from ≥ 3 phyla **320** (65.5) mL^{-1} 1,990 (284) mL^{-1} **757** (97.0) mL^{-1} ≥ 5 species from ≥ 3 phyla	≥ 1,000 mL^{-1} NR 1,700 (1,300) mL^{-1} 2,350 (1,390) mL^{-1} **994** (466) mL^{-1} NR
GSI **Uptake** **BE 1** **BE 2** **BE 3**	≥ 100,000 m^{-3} ≥ 5 species from ≥ 3 phyla 341,000 m^{-3} 353,000 m^{-3} 396,000 m^{-3} > 6 species from 3 phyla	≥ 1,000 mL^{-1} ≥ 5 species from ≥ 3 phyla 1,530/**605** mL^{-1} [D, E] 1,490/**572** mL^{-1} 2,340/**560** mL^{-1} > 5 species from > 3 phyla	≥ 1,000 mL^{-1} NR 10,000 (3,540) mL^{-1} 190,000 (28,300) mL^{-1} 86,700 (30,600) mL^{-1} NR

[A]Values shown in bold font and in shaded cells indicate the ETV or TQAP requirement was not met; all results in the table are from the TFs' Executive Summaries (Golden Bear Facility, 2012 and Great Ships Initiative, 2012). [B]Results are from the epifluorescent staining procedure (EPA, 2012). [C]Results from GBF are from counts on culture media (EPA, 2012), and results from GSI are from IDEXX SimPlates® (see Section 6.3.3 "Results"). [D]Size determined following the GSI departure from specification of any dimension of the entity ≥ 10 µm and individual cells < 50 µm. [E]Size determined by "maximum dimension on the smallest axis" (EPA, 2010). BE = biological efficacy, GBF = Golden Bear Facility, GSI = Great Ships Initiative, NR = no requirement, SD = 1 standard deviation, and TF = test facility.

Both TFs met their TQAP requirements for organisms ≥ 50 µm in all three tests, noting that GBF had received a DFS so a concentration of 10,000 living organisms m^{-3}, rather than 100,000 m^{-3}, was sufficient. The size class for organisms ≥ 10 µm and < 50 µm, however, presented a larger hurdle for both TFs: GBF met the requirement only in Test 2, and GSI met it in all three tests when sizing organisms according to the DFS for this project, that is, that measurements ≥ 10 µm and < 50 µm on *any* axis were acceptable. When sizing organisms the 'ETV' way, that is by "maximum dimension on the smallest axis" (EPA, 2010), the requirement was met in none of the GSI tests. In considering diversity requirements for the two largest size classes, they were met by both TFs in all tests. Finally, the requirement for culturable, aerobic heterotrophic bacteria was met in two of three instances at GBF (the mean concentration was slightly below 1,000 mL^{-1} in GBF Test 3); the GSI concentrations were 10- to 190-fold higher than the requirement.

Evaluating the uptake data from both TFs as a percentage of the ETV Protocol requirements, and *assuming no DFS for either facility*, differences are evident in the challenge water communities (Table 10). GBF used ambient populations with no augmentation, whereas GSI used ambient populations with augmentation of organisms in the ≥ 10 and < 50 µm size class. At both facilities, the middle size class (≥ 10 µm and < 50 µm) was difficult to meet; in only one test (BE 2 at GBF) was the requirement fulfilled. GSI exceeded the requirements for the other two size classes in all tests: surpassing the ≥ 50 µm size class by three- to four-fold and the < 10 µm size class by ten- to nearly two-hundred fold. GBF did not meet the minimum concentrations for organisms > 50 µm in all three trials and reached between approximately 20% – 90% of the requirement. Regarding organisms < 10 µm, GBF exceeded the requirement by approximately two-fold in BE Tests 1 and 2, and very nearly met the conditions for BE Test 3.

Table 10. Living organisms in the uptake water at test facilities compared to the requirements in the Environmental Technology Verification program Protocol.

Living organisms (method, where applicable)	Minimum ETV requirement	GBF organism concentrations (% of ETV requirement)[A]	GSI organism concentrations (% of ETV requirement)
≥ 50 µm organisms m^{-3}	100,000 m^{-3}	BE 1: **86,300 (86%)** BE 2: **18,900 (19%)** BE 3: **37,100 (37%)**	BE 1: 341,000 (341%) BE 2: 353,000 (353%) BE 3: 396,000 (396%)
≥ 10 and < 50 µm organisms mL^{-1} (epifluorescent microscopy)	1,000 mL^{-1}	BE 1: **320 (32%)** BE 2: 1,990 (199%) BE 3: **757 (76%)**	BE 1: **605 (60%)** BE 2: **572 (57%)** BE 3: **560 (56%)**
< 10 µm organisms mL^{-1} [B] (HPC for GBF, IDEXX® plates for GSI)	1,000 mL^{-1}	BE 1: 1,700 (170%) BE 2: 2,350 (235%) BE 3: **994 (99%)**	BE 1: 10,000 (1,000%) BE 2: 190,000 (19,000%) BE 3: 86,700 (8,670%)

[A]Values shown in bold font and in shaded cells indicate the ETV requirement was not met (i.e., the value is < 100%).
[B]Results from GBF are from counts on culture media, and results from GSI are from IDEXX SimPlates®. ETV = Environmental Technology Verification, GBF = Golden Bear Facility, GSI = Great Ships Initiative, HPC = heterotrophic plate counts.

6.1.3 Ballast Water Management System Flow Rates and Volumes

The BWMS is designed to operate at water flows up to 250 m^3 h^{-1} at a minimum pressure of 2.5 bar (37 psi) and a maximum pressure of 10 bar (150 psi). The ETV Protocol requires a total volume of 200 m^3 in each tank (treatment and control). Operating flow at each TF was provided by the main pump, and both GBF and GSI simultaneously filled the control and treatment tanks. GBF obtained source water from the sea chest of the vessel, while GSI obtained its water from a pipe extending out into Duluth-Superior Harbor. Both GBF

and GSI planned to test the system at a nominal 200 $m^3 h^{-1}$ flow, with GBF filling the treatment tank to a volume of 400 m^3 and GSI filling the treatment tank to 200 m^3.

The test plans from both GBF and GSI specified test parameters within the manufacturer's ratings and accounted for the fact the BWMS filtration system interrupts flow during backflush cycles. However, with the average treatment flow dependent on the frequency of backflushes, and this in turn dependent on source water conditions, it was impossible to specify a target treatment flow rate that resulted in an average flow of 200 $m^3 h^{-1}$. Additionally, at GBF, the ballast pump operated near maximum capacity as the BWMS approached the pressure required to initiate backflush, perhaps due to the position of the BWMS five decks above the water intake, and it was unknown if there was sufficient pump capacity to achieve an average flow of 200 $m^3 h^{-1}$ under full challenge conditions. The stress test during commissioning at 250 $m^3 h^{-1}$ flow was not conducted under challenge conditions with high suspended solids (as was the case during the BE tests). Thus, the VO provided guidance to both TFs that the target treatment flow rate should be set exclusive of backflush cycles (see Section 7.2 "Additional Suggestions to Improve the Protocol" for more discussion on this topic). The TFs were also requested to identify a tolerance range for their flow settings. The resulting nominal ballast flow of 200 $m^3 h^{-1}$ was set during sea-to-sea mode prior to treatment testing (i.e., setting flow conditions pumping ballast from the source back to the source through the inactive BWMS), but during operational testing, the treatment flow was effectively reduced due to flow stoppage during backflush operations. Because of this, both TFs planned to reduce flow to the control tank to match the average flow to the treatment tank, with concurrence from the VO. The actual amount of reduction on the control path was varied as needed during each test. This need to adjust control flow was due to the variable nature of treatment flow; the frequency of backflushing was a function of water quality and therefore not entirely predictable.

Ballast uptake flows as defined by each TF's TQAP are provided in Table 11. The upper half of the table provides target operating conditions, while the lower half of the table provides average flow data for each of the test cycles, exclusive of backflush cycles on the treatment uptake, as calculated by each TF. The table shows that measured treatment flow exclusive of backflush matched the target for all tests at GBF, and it was low by 3 $m^3 h^{-1}$ on one test at GSI. The low value was likely due to high sediment loading during uptake, resulting in high number of backflush cycles occurring. This condition reduced the average flow rate slightly below the target range due to normal operation of the treatment system, and it was not attributed to facility operational issues.

At GBF, the average uptake control flows were in the range expected (Table 11). At GSI, two of the tests produced lower control flows than the TQAP target range. During the second test, which had a high number of backflush operations, the control flow was manually adjusted to compensate, but an error was made in the value entered, and the resulting flow was below the target range; after the correct value was entered, the resulting value allowed the same total volume of water to be collected in the treatment and control tanks. The reported control flow value was below the GSI TQAP requirement. The third GSI test also had a high number of backflush cycles and required manual adjustment of the control flow rate; the resulting value allowed the same total volume of water to be collected in the treatment and control tanks, but it resulted in a lower average flow than the target. The final two test cycles at GSI both ran noticeably longer (106 min and 91 min, respectively) than the first test cycle (72 min) due to the higher number of backflush operations.

Table 11. Uptake flow rates of the ballast water management system at the test facilities.

Uptake operation	GBF range or target – TQAP (m^3 h^{-1})	GSI range – TQAP (m^3 h^{-1})
Treatment average w/o BF	180-220	180-210
Treatment average	~160	140-170
Control average	~160	140-170
Test and uptake operation	**GBF – data (m^3 h^{-1})**	**GSI – data[A] (m^3 h^{-1})**
BE 1: Treatment w/o BF	201.7	197
BE 1: Control average	162.3	164
BE 2: Treatment w/o BF	202.1	**177**
BE 2: Control average	170.9	**108**
BE 3: Treatment w/o BF	202.7	184
BE 3: Control average	181.5	**126**

[A]Values shown in bold font and in shaded cells indicate the TQAP requirement was not met. BE = biological efficacy, BF = backflush, GBF = Golden Bear Facility, GSI = Great Ships Initiative, TQAP = test quality assurance plan, w/o = without.

In considering target operating pressures, GBF did not cite a target operating pressure but instead allowed the ballast pump control system to determine operational pressure. In the GBF TQAP, it was indicated that during backflush cycles, the automatic pump control would detect reduced flow and increase the pump speed. This practice effectively increased line pressure and assisted the BWMS during backflush cycles. Also of note for GBF is the location of the BWMS on the 01 deck, five decks above the ballast pump, which was located in the "shaft alley" on the ship (the area with the propeller shaft, which extends from the propeller to the engine room). This vertical separation created additional head that the ballast pump had to overcome, effectively reducing the pressure available at the inlet of the BWMS. Conversely, at the outlet of the BWMS, there was suction as the water moved down, effectively providing a negative pressure.

At GSI, a range of target operating pressures was provided for both uptake and discharge operations. The human-machine interface (HMI) controlled and monitored pressure in the ballast piping. During uptake, the inlet pressure at the BWMS was set to 2.4 - 3.4 bar (25 - 50 psi). During discharge, it was set to 2.1 - 2.8 bar (30 - 40 psi) because the filter component of the BWMS was not engaged.

Note that GBF used two discharge flow rates to accommodate draining of their tanks, which were true ballast tanks emptied by a bell mouth suction pipe. All but the final 12" (30.5 cm) of water was discharged at the target rate. However, to prevent entrainment of air and enhance the volume removed from the tanks, the final portion was discharged at a lower rate of 50 m^3 h^{-1}. In contrast, GSI had aboveground tanks with a conical base and baffle to minimize air entrapment during draining; these were drained at target flow rates until they were empty. Both facilities were able to maintain the discharge flow ranges cited in their respective TQAPs (Table 12).

Table 12. Discharge flow rates of the ballast water management system at the test facilities.

Discharge operation	GBF range – TQAP ($m^3 h^{-1}$)	GSI range – TQAP ($m^3 h^{-1}$)
Treatment Average	180-220	190-210
Control Average	180-220	190-210
Test and discharge operation	**GBF mean[A] (discharge/stripping)**	**GSI mean ($m^3 h^{-1}$)**
BE 1: Treatment	186.3 (202.9 / 56.0)	195
BE 1: Control	196.4 (202.1 / 85.6)	199
BE 2: Treatment	188.1 (203.4 / 50.4)	197
BE 2: Control	185.9 (204.4[B] / 51.4)	199
BE 3: Treatment	184.1 (199.0 / 59.8)	197
BE 3: Control	183.3 (198.6[B] / 53.4)	200

[A]In all instances, the TQAP requirement was met. [B]Data regarding test cycle operations in the GBF verification report are slightly different from those presented in the GBF Executive Summary, but the values are within 10% of each other; the values from the Executive Summary are reported here. BE = biological efficacy, GBF = Golden Bear Facility, GSI = Great Ships Initiative, and TQAP = test quality assurance plan.

6.2 Test Facility Physical Configuration

6.2.1 Ballast Water Management System Installation and Use

Installing the BWMS required each TF to provide several power and piping interface connections. Once the initial piping modifications to allow inflow and outflow through the same end of the container had been completed, power and piping were configured as necessary. These tasks were accomplished with relative ease at both facilities. The location for the BWMS installation, however, was unique to each facility. GBF is on a working ship and thus has both multiple levels and a compartmentalized structure, while the layout of GSI is on a wharf and is essentially planar or a single-level structure. The differences in the facilities' temporary installations of the BWMS and ballasting infrastructure are described in this section.

6.2.1.1 *Golden Bear Facility Installation of the Ballast Water Management System on the 01 Deck*

With the GBF installation of the BWMS on a working vessel, the system was hoisted onto (and later off) the vessel near the vicinity of the ship's crane. GBF provided tie downs, ballast piping access, and air and power connections on the 01 deck.

The ballast pipes connecting the BWMS to the ballast pumps and treatment tank extended five decks below to shaft alley, where the pump, valves, sea chest, and controls for the ballast system were located. The treatment tank was on the port (shipboard location reference = 3-154-2) and control tank on the starboard (shipboard location reference = 3-154-1). Both were located in the aft portion of the ship, directly above the shaft alley space, with the tank tops on the second deck (two decks below the 01 deck). The control and treatment tanks were symmetrical about the centerline of the ship, with capacities of 441 MT (port) and 432 MT (starboard). Each tank was ballasted through 8" (20.3 cm) piping terminated in a vertical bell mouth 2" (5.1 cm) above the tank bottom. Typically, the last 12" (30 cm) of water in the tank was drained at a reduced flow rate to minimize air entrapment and maintain suction (stripping the tank). This practice minimized the residual volume in the tank. The amount of water in each tank affected the position of the ship (trim and list), and these changed continuously with the tank levels.

6.2.1.2 Great Ships Initiative Installation in Laboratory Building 1

The Laboratory Building 1 at GSI can accommodate two BWMSs, although only the BWMS from the project occupied the building during Intercomparison testing. The building has a high overhead enclosure and concrete spill pads under the piping connections for the treatment systems. The containerized BWMS was installed on the gravel floor within the building, and temporary PVC piping was used to connect the inlet and outlet of the BWMS to the rigid steel piping of the facility, while the backflush line was connected with a flexible hose. Power and air connections were available within the building. The vertical separations for piping and system connections between ballast pump, sample collection ports, and control and treatment tanks were generally only a few feet (approximately 1 m), so no extra head capacity was required to reach the BWMS.

The control and treatment tanks at GSI were above ground, and the tanks had a conical base with the fill and drain piping connection at the lowest level in the tank. A baffle plate was located above the piping connection opening to minimize vortexing and subsequent air entrapment during drain operations. Typically, the tanks were drained at constant flow until they were completely empty. The tanks also contained internal single-blade mixers at the bottom to circulate the water and reduce stratification; the mixers were validated to determine they did not induce additional mortality to organisms (GSI, unpubl. data), and in this study, the concentration of living organisms in the control discharge samples was approximately 4,000- to 5,000-fold higher than the ETV requirement of 100 m^{-3}.

6.2.1.3 Comparison of the Two Installations

The ETV protocol provides no guidance on tank design or configuration for verification testing, and the two facilities used different approaches. GBF's control and treatment tanks employed multiple flow rates to maximize water removal, and a bell mouth suction system, commonly seen on vessels, was used to drain the tanks. Significant sediment loading was observed during tank stripping; this feature is likely typical of ballasting operations on a ship. The GSI facility, with aboveground tanks and a drain at the lowest point in the tank, provided complete drainage, and drained both sets of tanks at a constant flow rate over the entire drain operation. Also, while keeping the contents of the tank continuously mixed should reduce "patchiness" and seems desirable from a scientific standpoint, it was not necessarily representative of typical ship conditions.

The GBF installation of the BWMS had significant height differences from the ballast pump, which resulted in long piping lengths and additional head pressure. Based on casual review of BWMS literature from several vendors, typical installation locations are near the ballast pumps, often in the engine room. In this instance, the GSI installation with treatment system and ballast pumps having little vertical separation between sea level and the BWMS seems more conventional. The motion introduced at GBF during changing of tank contents, while encountered regularly by ships, was likely smaller than would be found on a vessel underway and likely did not affect the BWMS or the test outcome. These points notwithstanding, the goal of the ETV testing is to conduct rigorous, well-controlled tests, and thus, as the protocol is currently written, the degree to which ship conditions are replicated is not as important as the repeatability of the tests.

6.2.2 Control and Instrumentation

Both TFs implemented automated data logging (periodic recording of data values) for numerous test parameters, along with some degree of automated control (autonomous adjustment of variables such as valve position or flow rate to match desired operational set points). While automation of control functions is not an ETV requirement, a TF may choose to automate certain processes to maintain or improve control

of test parameters over long periods or to improve repeatability between tests. This section describes how each of the participating TFs implemented control processes and recorded data from instrumentation within the facility. It should be noted that both TFs plan to further upgrade their control and monitoring capabilities, so the discussion below pertains to observations during the Intercomparison Project.

6.2.2.1 *Golden Bear Facility Control and Instrumentation*

At GBF, most controls were operated manually with only the ballast pump capable of autonomous operation. This widespread use of manual control valves for test operations was described in part by GBF as a philosophical decision by faculty and crew, to ensure student involvement in testing as part of their educational experience on the ship. Further, the available budget and the time required to plan and implement automation within the requisite time constraints of this project were also factors. Just prior to (and in preparation for) the Intercomparison Project, GBF installed and implemented automated data logging and computer indicators to support manual control of ballast treatment test operations. The computer system (IMAC) passively monitored many test parameters and used operator input to confirm parameters not integrated into the control system (e.g., the position of valves without position sensors). Additionally, IMAC provided data entry and served as a communications system between the various areas of activity on the ship. Computers connected to IMAC showed the status of operations, and in this manner, computers used in shaft alley, in the information technology (IT) space, or on the main deck provided visual, real-time updates on testing. The multiple display views on IMAC showed the status of the ballast system in real time, and it showed parameters being logged, such as flows, tank levels, pressures, temperatures, etc. This widespread accessibility to real-time data is advantageous during testing.

As part of the installation of the BWMS on the T/S *Golden Bear*, control signals were wired between the BWMS control panel and the IMAC. This allowed the GBF Lead Operator to not only control the ballast pump, but also to start and stop the BWMS under computer control. It also allowed the alarm signal from the BWMS to generate an alert on the IMAC display, although no additional diagnostics were available without accessing the BWMS controls. The IMAC software provided both data entry screens and diagrams (such as piping diagrams with valve configurations) to control and log system operations. The IMAC software package incorporated a communications protocol developed by GBF called the Ballast Order Telegraph (BOT) to communicate and confirm system status between the operations and sampling teams. Its design was based on the Engine Order Telegraph log used on ships to communicate between the navigation bridge and the engine room. The IMAC display on computer terminal screens had “buttons” to change an operational state and initiate digital record keeping during testing. In transitioning from one operating state to another, the IMAC used truth tables to ensure the correct parameters (e.g., valve lineups) were in place prior to starting the transition.

Sampling operations were controlled manually and continuously monitored (and adjusted as necessary) by the scientific staff during uptake and discharge. This task included not only concentration of organisms (using plankton nets), but also maintaining the flows entering all of the sample collection tubs and whole water sample containers. To ensure the flow rates were maintained as planned, a table of the target flow values for the uptake and discharge sequences was posted at the sampling station. Sample flows to the plankton nets were instrumented with flow sensors, so these values were continuously logged by IMAC. When collecting samples during uptake operations, every time the GFE went into backflush mode, an indicator on the IMAC display screen would cue an operator to manually close the sample collection valve at the main ballast line and to re-open it when the BWMS resumed filtering. This procedure prevented sampling of the ballast stream during non-operational times, but more importantly, it prevented the purging (i.e., air infiltration) of the sample collection apparatus during the no-flow conditions. If the sampling

system had not been isolated during backflush operations, it could have taken several minutes for flow conditions to re-stabilize. These events can be observed with one-minute resolution in the sample flow data logged by IMAC.

GBF also had installed several in-line sensors, providing a low-volume side stream of uptake water to a suite of water chemistry sensors. This side stream was driven from the ballast pump pressure and drained into the bilge. These sensors were located in shaft alley and continuously monitored parameters such as salinity, optical transmission, conductivity, and temperature. These data were also logged by the computer system.

Because IMAC was installed immediately prior to the Intercomparison Project and used for the first time during these BE tests, some difficulties were encountered. Specifically, the system generated substantial amounts of data, but the project timeline did not permit the development of tools to organize, analyze, and generate reports of the data; this development began after testing was completed. Thus, processing, auditing, analyzing, and subsequently reporting the operational data was a lengthy process.

6.2.2.2 *Great Ships Initiative Control and Instrumentation*

At GSI, a computer-based control and data acquisition system was used, which was referred to as the HMI. The system included computer control of pumps for ballasting, water circulation, biological and water quality augmentation, and retention tank mixers, as well as flow control valves for both ballasting and sampling collections. Similar to the GBF system, the operator display provided various screen displays of piping diagrams with real-time operational values, and it provided for data entry and system checks when transitioning from one operational state to another. The HMI automated several functions, such as maintaining correct ballast and sample flows during uptake and discharge operations. At GSI, the HMI was accessed by only one operator and was situated in one location (the control room).

GSI chose not to interface control or alarm signals from the BWMS system to the facility's HMI, with the exception of a status signal to indicate the BWMS had entered backflush mode. Instead, GSI engineers preferred to use the remote panel interface supplied with the treatment system to manually operate and monitor the remote control panel for the BWMS's status. Because the BWMS and the control room were both in Laboratory Building 1 and less than 50' (15 m) apart, this task was achieved by running a cable between the remote panel interface and the BWMS. This decision was driven largely to divide the management of the BWMS and the facility between the two engineers at GSI. It was also partly due to limited free space on the HMI graphical display; there was really no room to fit the information needed to monitor and control the BWMS on the operator display in addition to all of the facility-specific control and monitoring information, and a remote display panel had been provided with the BWMS. Another factor in this decision was GSI's preference to not impose BWMS-specific custom programming on their in-house control system software.

To temporarily stop sampling during backflush operations, the HMI automatically paused sample collection when it detected that the outlet butterfly valve on the BWMS was completely closed. Representatives from Rockwell Automation, who programed and maintained the GSI HMI system, tied the butterfly valve status signal (available as an output from the BWMS) into the GSI control system and included programming to monitor and automatically stop sampling during backflush operations. This interlock was successfully tested by GSI prior to running any water automatically and worked without incident throughout the testing.

During testing, the Lead Operator was stationed at the HMI controls with the remote control panel for the BWMS nearby. Readouts from the BWMS were manually logged by the Lead Operator, while TF operating parameters were automatically logged by the HMI. Essentially all functions for TF operations were controlled from the HMI control system, along with the collection of sample water for organisms ≥ 50 µm. This arrangement allowed the other members of the operations staff to observe (and attend to as needed) the various pieces of equipment, such as phytoplankton holding ponds and water chemistry augmentation tanks, BWMS, and sampling stations. Water chemistry measurements were collected using stand-alone, hand held instruments by scientific staff, sampling staff, or both. Data from these instruments were independently uploaded into the facility database, and were not part of the HMI. The facility has since added in-line sensors that are configured with the facility to collect accurate measurements.

Prior to the Intercomparison Project, GSI had not used their HMI to perform full system testing, and few tools for data analysis and report generation for operational data were available at the start of the project. Data analyses were performed by electronically reading data logs into spreadsheets and generating plots one at a time. The TF, however, did have a mature IT capability for storing and organizing data.

6.2.3 Living Organism and Water Quality Augmentation

To meet the challenge water requirements identified in the ETV Protocol, TFs may augment the naturally occurring concentrations of organisms and inorganic dissolved and particulate constituents. One means of augmenting the living organisms is by adding "standard test organisms" (STOs) to challenge water, which are defined as "biological organisms of known types and abundance that have been previously evaluated for their level of resistance to physical and/or chemical stressors representing ballast water technology" (EPA, 2010). STOs may be a native or non-native species (or multiple species) cultured in the laboratory in high densities or concentrated from natural waters, but importantly, the selected STO(s) must have been assessed under various scenarios representing treatment by a BWMS so the level of resistance or hardiness to treatment is known. Neither TF used STOs.

At GBF, no augmentation equipment or procedures were in place at the start of the project, and the means to augment natural waters could not be realistically constructed and validated in the project timeline. Therefore, un-augmented water from Carquinez Strait was used in all BE tests at GBF, and DFSs of some constituents outside the regular ETV Protocol-specified ranges were agreed (Table 2). At GSI, equipment and procedures for augmenting organisms (by concentrating organisms from nearby water) and water quality constituents were in place and were used during the BE tests (Table 13). While the means to collect and inject organisms was validated prior to the Intercomparison Project, to the VO's knowledge, the effect of adding MM and POC to challenge water was not validated, although GSI was aware the methods have been used by the VO with no ill effects.

Table 13. Approaches to augmenting living organisms and water quality parameters according to the Environmental Technology Verification Protocol and by the Great Ships Initiative.

Protocol or TF	Biological and water quality parameters and caveats			
	Living organisms (in any size class)	**Particulate organic matter**	**Dissolved organic matter**	**Mineral matter**
ETV	The means to concentrate organisms for augmentation is not directly addressed; injection of organisms into challenge water must minimize mortality to the extent possible, result in a well-mixed distribution, and be validated by the TF	Addition of humic material (e.g., Micromate humates from Mesa Verde Resources); TF should verify that substance does not inhibit or stimulate ambient or test organisms, and the injection should be upstream from WQ sampling to represent a well-mixed sample	Addition of *Camellia sinesis*, decaffeinated iced tea mix; TF should address the effect of additives on organisms and should verify that substance does not inhibit or stimulate ambient organisms; the injection should be upstream from WQ sampling to represent a well-mixed sample	Addition of commercially available clay minerals (majority of particles ≤ 10 µm and < 50 µm, e.g., ISO Medium and Coarse Test Dust); TF should maintain sediment in homogenous suspension prior to adding it to challenge water and verify that substance does not inhibit or stimulate ambient organisms; the injection should be upstream from WQ sampling to represent a well-mixed sample
GBF	No augmentation	No augmentation	No augmentation	No augmentation
GSI	Conducted plankton tows in Duluth-Superior Harbor for organisms ≥ 10 µm and < 50 µm; kept organisms in aerated holding ponds for ≤ 48 h and added them to challenge water at a constant rate with a diaphragm pump; process validated by GSI to demonstrate survival of organisms and equality in control and treatment tracks	Added humic material (i.e., Micromate) specified in the ETV Protocol at a constant rate using a peristaltic pump; GSI sterilized the Micromate prior to testing to ensure no living organisms were inadvertently added to the challenge water (the ETV Protocol does not require this step)	No augmentation	Added test dust specified in ETV Protocol at a constant rate using a peristaltic pump; although the Protocol did not specify it, the test dust was sterilized by GSI prior to tests to ensure no living organisms were inadvertently added to the challenge water

ETV= Environmental Technology Verification program, GSI= Great Ships Initiative, TF = test facility, and WQ = water quality.

6.2.4 Sampling Methodology

The two TFs used different approaches for sampling organisms, which are described below.

6.2.4.1 Sampling Methodology — Golden Bear Facility

At GBF, in-line sample ports were installed in the center of the flow stream to collect time-integrated samples. Because the sample volumes collected during the uptake and control and treatment discharge operations differed, two sample ports were used to ensure samples were collected in a sub-isokinetic manner (Figure 5). They were sized following the guidance in the ETV Protocol Section 5.3.2.5 "Design of In-line Sampling Apparatus" (EPA, 2010), using 1.25" (3.18 cm) for the uptake and control tank discharge samples and 1.5" (3.81 cm) for the treatment discharge samples. The sample ports were located in the center of the flow stream. The VO evaluated the size of the sample ports during the review of GBF's TQAP and concurred they were sized according to the ETV Protocol.

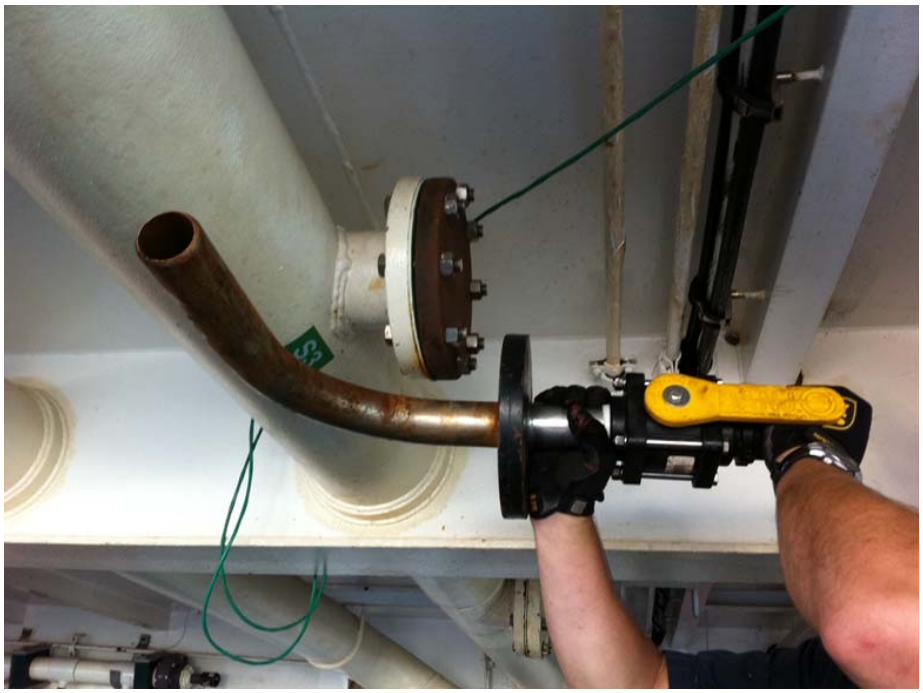

Figure 5. Sample port used at the Golden Bear Facility. This photograph shows the 4.09 cm (1.61"; measurements are the inside diameter) sample port used for uptake and control discharge sampling operations.

To collect organisms ≥ 50 µm, GBF used a flow-through sampling apparatus with two independent sampling arrays, each consisting of three individual sample tubs. The flow from the main sample port was split into each of the sample tubs, so GBF obtained a single sub-sample of the main ballast flow that was split four ways (into three tubs and one waste stream). For uptake sampling only, one array of tubs was used, whereas both arrays of sample tubs were used for control and treatment discharge sampling. For discharge, first the treatment discharge was sampled (in one array), and then the control discharge was sampled (in the second array). A plankton net with 35 µm mesh (with an un-recorded mouth diameter and 40 cm length) was suspended inside each of the sample tubs. The water was directed from the sample port with a hose to the inlet of the sampling manifold above the sampling tubs (Figure 6). After the inlet, the water was mixed in the three-way manifold before being directed through clear hoses to each of the

individual sample tubs, which each contained a plankton net. Sample flow streams were directed into the plankton nets below the surface of the water. Time-averaged sample flow from the sample port was achieved by manually throttling a dump valve downstream of the 3-way manifold (using a data table developed by GBF). Each clear hose from the manifold to a sample tank was connected to a flow control pipe consisting of a flow meter and a manual flow control valve. The sample flow to each tub was controlled manually by a member of the GBF science team using the flow control valve. The water level within the sample tubs was set using standpipes on the exterior of the tubs.

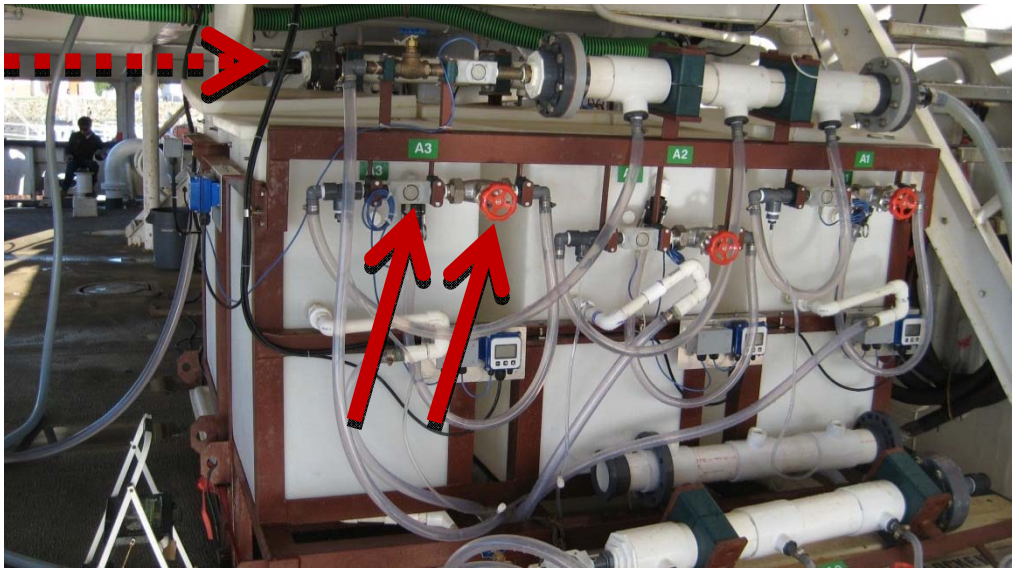

Figure 6. Sampling apparatus used by the Golden Bear Facility to concentrate organisms ≥ 50 µm in plankton nets (inside the white tubs; not visible). The dashed, red arrow indicates the inlet to the sampling manifold from the sample port. The solid, red arrows indicate the flow meter (left arrow) and red manual flow control valve (right arrow) used to control sample flow rate to the individual white tubs.

For all biological samples for organisms < 50 µm, GBF collected whole-water samples using a pinch valve connected to a tee upstream of the sample tubs (Figure 7). A pinch valve was connected to the clear hose that controlled the flow rate. GBF used this setup to collect three samples. For each sample, the clear hose was placed into one of three 20 L plastic carboys. The carboys were used to collect samples for both the ≥10 µm and < 50 µm as well as the < 10 µm size classes. Each carboy was well mixed, and then two 1 L samples were then taken from each carboy.

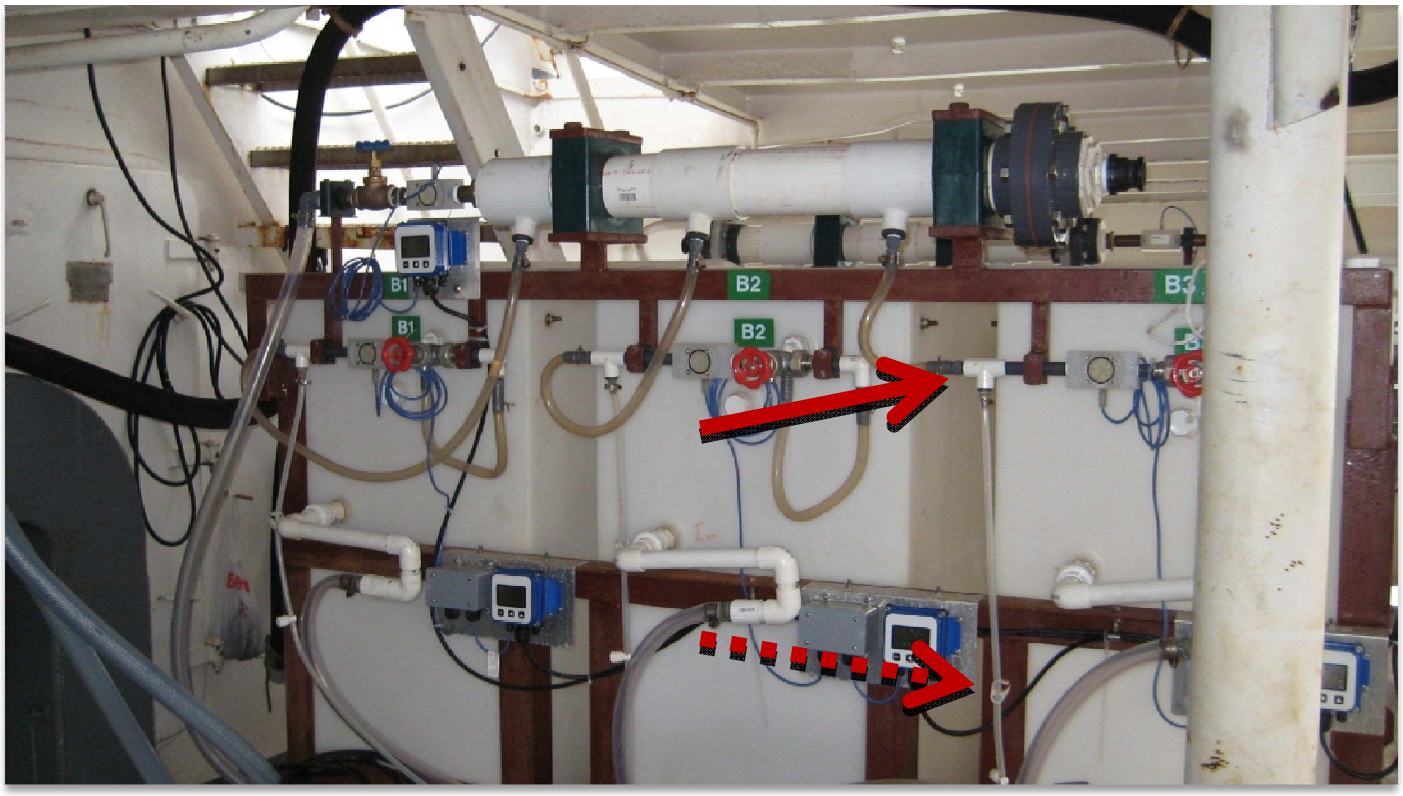

Figure 7. Sample collection tee (marked by the solid, red arrow) used by the Golden Bear Facility to collect organisms < 50 µm (≥ 10 µm and < 50 µm as well as organisms < 10 µm). The dashed, red arrow indicates the pinch valve used to control the sample flow.

6.2.4.2 Sampling Methodology — Great Ships Initiative

GSI used up to three 3.8 cm (1.5 in) diameter, in-line sample ports to obtain time-averaged biological samples from the center of the ballast water flow stream for each sampling event (Figure 8; validated by GSI, unpubl. data). Each sample port was connected to the bottom of one of the 3.8 m^3 biological sample collection tubs (Figure 9), such that each tub represented a sub-sample of the main ballast flow.

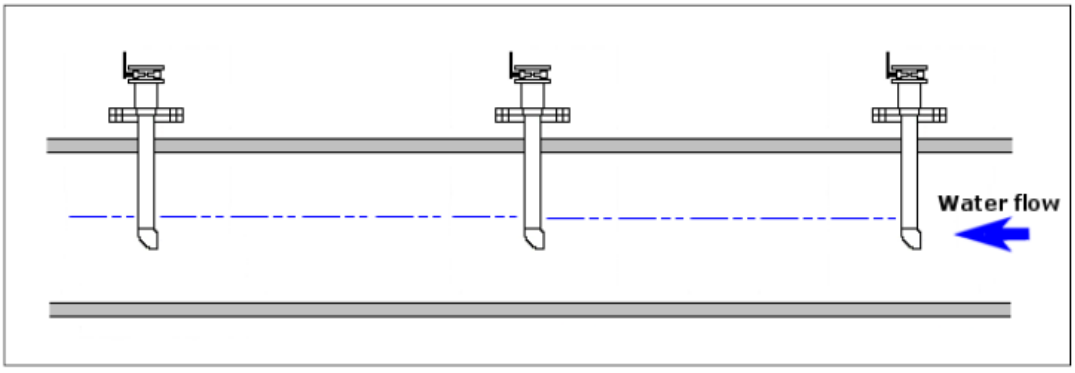

Figure 8. Sample ports used by the Great Ships Initiative to collect organisms ≥ 50 µm (Great Ships Initiative, 2012).

As was the case at GBF, the sample ports were located in the center of the main ballast flow stream and were designed per the guidance in the ETV Protocol Section 5.3.2.5 "Design of In-line Sampling Apparatus" (EPA, 2010). The VO evaluated the sample port sizes during review of this facility's initial TQAP and concurred that the sample ports were sized per the ETV Protocol guidelines.

GSI used 1.5 in (3.8 cm) diameter sample ports to collect organisms ≥ 50 µm. Samples were collected in the large white sample collection tubs (Figure 9). The samples were collected over the entire uptake or discharge operation, and flow control for the sample water was automatically controlled using a diaphragm valve linked to a flow sensor. The volume and flow rate for an event for each sample tub was recorded every 5 seconds by the facility control system.

Figure 9. Biological sample collection tubs at the Great Ships Initiative; tubs are white with conical bottoms.

After an uptake or discharge operation was completed, samples for organisms ≥ 50 µm were collected sequentially from the individual sample collection tubs using 35 µm mesh plankton nets (30 cm mouth diameter and 90 cm long). The nets were hung below the drain spout of the tub, with the net suspended in a 130 L plastic barrel, and the flow was directed into the nets using a hose that terminated in the bottom of the plankton net. The nets were fitted with 1 L cod ends (one plankton net was used for each sample collection tub). The entire contents of the sample collection tub were drained through the plankton nets. Samples were analzed immediately after collection in a sequential process (i.e., one sample collection tub at a time was drained, processed, and analyzed).

GSI collected time-averaged whole-water samples for organisms < 50 µm using "seep" sampling. A side stream of water was directed from the sample port line through a tee that was connected to a flow meter, clear hose, and pinch valve. The flow rate was manually controlled by an operator who monitored the flow using the visual flow meter (Figure 10). The clear hose was directed to a 19 L carboy. Each of the three sample ports in the discharge operation had seep samplers installed, and three sets of samples were obtained for all uptake and discharge operations for this project. The whole water samples collected with this method were used to enumerate organisms in the two smallest size classes. Each of the 19 L carboys was well mixed, and 1 L was removed for organisms < 10 µm, and 1 L was removed for organisms ≥ 10 µm and < 50 µm. Thus, 2 L was removed from each 19 L carboy and analyzed.

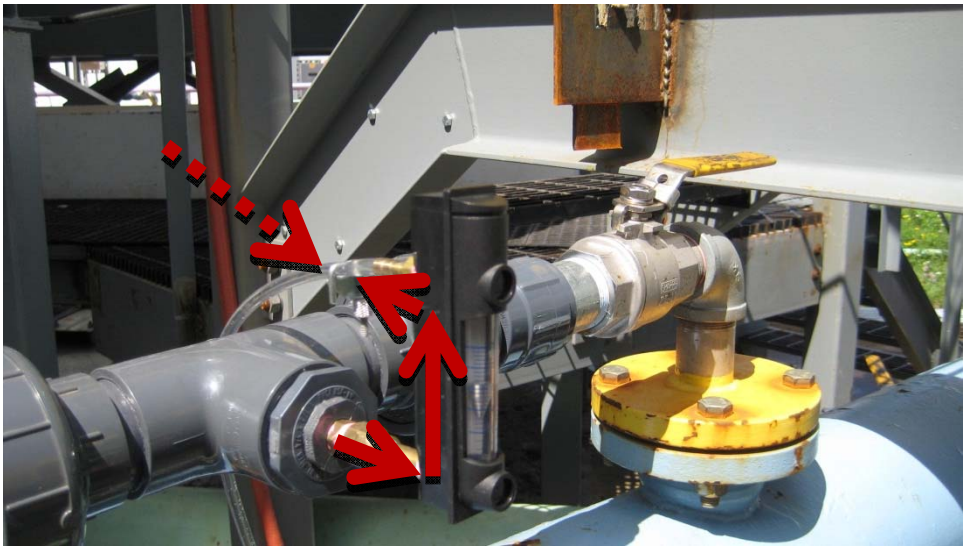

Figure 10. "Seep" sampling device at the Great Ships Initiative for collecting organisms < 50 µm. The solid, red arrows depict the sample flow from the main line, through the flow meter, to plastic tubing, which leads to a carboy (not shown). The dashed, red arrow indicates the pinch valve on the plastic tubing, which controls flow to the carboy.

6.2.4.3 *Sampling Methodology — Statistical Analysis of Variance and Partitioning between Test Facilities*

The GBF and GSI facilities collected sample water differently during the ballast discharge process: GBF used a single port off the ballast water discharge, from which water was split into three sample tubs, whereas GSI used up to three separate ports, with water from each port transferred to its own individual tub (Figure 11). Therefore, both TFs SOPs required collection of three subsamples.

Figure 11. The three sample ports used for biological sampling at the Great Ships Initiative (left image) and one sample port used at the Golden Bear Facility (right image). The red arrows identify the sample ports in the left image.

While GSI used multiple sample ports, which were separated by ten pipe diameters. Ten pipe diameters are typically used to ensure that the flow is well mixed and a uniform flow profile across the pipe has resumed, i.e. that the previous intrusion has no perceptible impact on the mixing or velocity profile. There were control mechanisms in place to ensure that the sample was homogeneous at all sampling points. That is, water was discharged sequentially from the treatment and control tanks, while flow control valves and system logic ensured that flow rates were equivalent and proportional, with a well-mixed flow target of 200 $m^3\ h^{-1}$ (880 gal min^{-1} per track). Similarly, GBF used a flow rate target of 200 $m^3\ h^{-1}$ (880 gal min^{-1}) as well

as passive monitoring of valve positions, flow rate, pressure, temperature, water chemistry, and sample flow rate to ensure sampling adhered to target parameters. Additionally, at GSI, samples obtained at the three ports were verified as independent (i.e., sample collected from downstream ports were not affected by the presence of upstream ports (GSI, unpubl. data).

Given the difference between the sampling approaches at the TFs, it is useful to consider any possible effects on the estimates of organisms, specifically, the issue of estimated variance from the sample. As a measure of dispersion, there is naturally expected variation that would occur in any random sample of a population while the population mean and variance remain unknown (Johnson and Bhattacharyya, 2006). For sample collection and testing of *treated* water, with relatively low concentrations of organisms, it is necessary to operate under the same assumptions as the Poisson distribution:

1. The sample flow is well-mixed, and
2. organisms within the flow are randomly distributed and independent (Lemieux et al., 2008; Richard et al., 2008; Lee et al., 2010).

If the aforementioned distribution assumptions are met (and assuming no loss occurs during sampling handling and processing), the variance in the subsamples (collected at GBF and GSI) should not be any greater other than the natural expected variance occurring in any given random sample. The same criteria apply to both TFs: whether the sample is taken from a single sample port or multiple ports, it must maintain a well-mixed flow to ensure that the water is unmodified between sample points. This condition will allow for appropriate random sampling and the use of Poisson methodology for statistical testing and verification for rare populations, for example, treated ballast water. As long as these criteria are fulfilled, the difference in the TFs' sample ports will have no impact on the analysis.

The possibility also exists that chain-forming organisms might exist in the sample water. These organisms have the potential to violate the assumption of independence of the random sample (the second criterion listed above), as one chain-forming organism's location would, in fact, affect the location of other organisms in its chain. It is acknowledged that this is a potential issue in the analysis; however, methods used in comparison of means are known to frequently work well even while violating assumptions, particularly with large sample sizes and if the violations are known to not be critical (Whitlock and Schluter, 2009). Addressing the problems posed by chain-forming organisms is a potential area for future study.

The assumptions above are true for non-sparse ($n > 30$) samples as well. With an adequately large sample size, a well-mixed, random sample will allow common statistical methods to be used, both parametric and non-parametric (Whitlock and Schluter, 2009). These statistical approaches to non-sparse samples are also appropriate for examining samples from challenge water, and for the purposes of this Intercomparison Project, samples from the treated water were analyzed as non-sparse samples, because the concentration of living organisms ≥ 50 µm was greater than 30.

6.2.4.4 *Sampling Results for Organisms ≥ 50 µm*

Both TFs collected sample volumes of at least 2.6 m^3 on uptake for all tests (Table 14) in accordance with guidance suggested in the ETV Protocol. For control tank discharges, the TFs collected at least 3 m^3 on nearly all BE tests, with the exception of BE test 1 by GBF, in which 2.9 m^3 was collected. For both uptake and discharge operations from the control tank, both TFs assumed the concentrations of organisms would be high enough (≥ 100 organisms m^{-3}) to assume a normal distribution in the statistical analysis of the data collected.

Table 14. Uptake and discharge sample volumes and volumetric flow rates for organisms ≥ 50 µm at the test facilities.

TF Operation	Sample volume (m^3)	Uptake or discharge sample flow rate ($m^3\ h^{-1}$)	Sample port size cm (inches)[A]	Did the sample port size meet ETV Protocol Requirements? (Y/N)
Uptake Sample Volumes and Sample Flow Rates ≥ 50 µm organisms				
GBF Uptake[B]	BE 1: 3.6 BE 2: 3.7 BE 3: 3.6	BE 1: 1.1 BE 2: 0.9 BE 3: 1.5	3.18 (1.25)	Y
GSI Control tank uptake	BE 1: 2.6 BE 2: 3.1 BE 3: 3.2	BE 1: 2.3 BE 2: 1.8 BE 3: 2.1	3.8 (1.5)	Y
GSI Treatment tank uptake	BE 1: 3.0 BE 2: 2.9 BE 3: 3.0	BE 1: 3.0 BE 2: 2.7 BE 3: 2.9	3.8 (1.5)	Y
Discharge Sample Volumes and Sample Flow Rates ≥ 50 µm organisms				
GBF Control tank discharge	BE 1: 2.9 BE 2: 3.3 BE 3: 3.2	BE 1: 2.5 BE 2: 1.6 BE 3: 1.9	3.2 (1.25)	Y
GBF Treatment tank discharge	BE 1: 9.6 BE 2: 9.3 BE 3: 9.5	BE 1: 4.6 BE 2: 3.0 BE 3: 4.1	3.8 (1.5)	Y
GSI Control tank discharge	BE 1: 3.7 BE 2: 3.6 BE 3: 3.7	BE 1: 3.7 BE 2: 3.7 BE 3: 3.6	3.8 (1.5)	Y
GSI Treatment tank discharge[C]	BE 1: 3.7 BE 2: 3.4 BE 3: 3.6	BE 1: 3.7 BE 2: 3.7 BE 3: 3.7	3.8 (1.5)	Y

[A]All sample ports were L-shaped pipes inserted into the center of the main ballast flow stream and sized according to EPA, 2010. [B]GBF processed only one sample on uptake because validation tests showed samples were equivalent after the main ballast water split to the treatment and control tracks. [C]GSI collected two sets of time-averaged samples on treatment tank discharge but analyzed only one set of samples because the density of organisms was higher than expected. BE = biological efficacy, ETV = Environmental Technology Verification program, GBF = Golden Bear Facility, GSI = Great Ships Initiative, and Y = yes.

Due to the configuration of the vessel, the GBF tank volumes were twice as large as required by the ETV Protocol (400 m^3). This led to a difference in the sample volumes collected by the TFs for treatment tank discharge operations for organisms ≥ 50 µm. GBF collected over 9 m^3 in all three BE tests, and used a 1.5" (3.81 cm) sample port rather than the 1.25" (3.18 cm) sample port used on uptake and discharge from the control tank. During development of its TQAP, GBF assumed that the treatment system would result in very low concentrations of living organisms in the treatment discharge water. The larger sample port was needed to ensure the larger volume of sample collected still met the ETV Protocol guidelines for sample port size.

GSI, on the other hand, collected three sample tubs of water for ≥ 50 µm sample analysis on treatment discharge. The sample volumes ranged from 3.4 m^3 to 3.7 m^3. In GSI's TQAP, it was specified the entire contents of the first tub would be concentrated through a plankton net, and then the entire concentrated volume of a given tub would be analyzed prior to draining the next tub (assuming the tubs contained sparse concentration of organisms, warranting analysis of all of the sample volume). After the initial analysis was

completed for each BE test, GSI determined the samples had high concentrations (rather than the expected rare populations) and they would not need to concentrate and analyze the organisms collected in the third sample tub (i.e., the organisms in two sample tubs were concentrated and analyzed for all three BE tests). This conclusion was based on analysis of the first tub resulting in at least 150 live organisms in a volume corresponding to less than 8.0 L of the original sample, which was extrapolated to a density of more than 10,000 live organisms m^{-3}.

6.2.4.5 Sampling Results for Organisms ≥ 10 µm and < 50µm and Organisms <10 µm

Both GSI and GBF collected time-averaged whole water samples used to analyze organisms in the two smallest size classes (≥ 10 µm and < 50 µm, < 10 µm; Table 15). The TFs collected one (GSI uptake and control discharge) or three (all GBF operations and GSI treatment discharge) subsamples from tubing connected to a sample port in a sampling manifold (GBF) or the main ballast line (GSI). The samples for the two smallest size classes were essentially drip samples obtained via a side port or tee fitting and were collected in 19 L (GSI) or 20 L (GBF) carboys. The ETV Protocol is silent on the methods for collecting low-flow, small-volume volume samples for biological analyses. Given that the flow in the sample line connected to the sample port may or may not have been fully developed turbulent flow where subsampled, the method used was deemed acceptable by the VO. Both TFs mixed the contents of the carboys to ensure homogeneity and then poured the water from the carboys into 1 L containers for processing both size classes of organisms.

Table 15. Uptake and discharge sample volumes and volumetric flow rates for organisms < 50 µm (≥ 10 µm and < 50 µm; < 10 µm) at the test facilities.

TF operation	Total sample volume (L)	Uptake or discharge sample flow rate[A]	Sample port size inches (cm)[B]
GBF Uptake	BE 1: 60 BE 2: 60 BE 3: 60	NM-Continuous	0.125 (0.32)
GSI Control tank uptake	BE 1: 15.0 BE 2: 19.0 BE 3: 15.0	NM-Continuous	0.125 (0.32)
GSI Treatment tank uptake	BE 1: 15.0 BE 2: 19.5[C] BE 3: 16.0	NM-Continuous	0.125 (0.32)
GBF Discharge	BE 1: 60 BE 2: 60 BE 3: 60	NM-Continuous	0.125 (0.32)
GSI Control tank discharge	BE 1: 15.0 BE 2: 14.5 BE 3: 15.5	NM-Continuous	0.125 (0.32)
GSI Treatment tank discharge	BE 1: 46.5 BE 2: 42.0 BE 3: 46.5	NM-Continuous	0.125 (0.32)

[A]Neither facility automatically recorded flow rate for this measurement; instead, the flow rate was visually monitored by test operators to ensure a continuous flow rate was achieved. [B]Sample ports for this size class at GBF and GSI were tees. [C]The volume was estimated, as the top gradation of the carboy (19 L) was exceeded, while the total volume of the carboy (20 L) was not exceeded. BE = biological efficacy, GBF = Golden Bear Facility, GSI = Great Ships Initiative, NM = not measured, and Y = yes.

6.3 Verification Testing

6.3.1 Commissioning

The ETV Protocol describes commissioning as a set of tests and start-up requirements to validate that the BWMS is installed correctly within the TF. Commissioning allowed each facility to install the BWMS and ensure it could reach stable operating conditions. As part of the commissioning process, the BWMS manufacturer provided on-site technical support to each TF for one week to ensure the BWMS was installed and setup properly. The BWMS technician also provided training to the TF staff to ensure they were able to operate the equipment independent of the BWMS manufacturer. The ETV Protocol recommends the TF to conduct one valid, full-scale BE test, independent of the BWMS vendor that meets all requisite challenge conditions as a means to fully exercise the TF operations under test conditions and demonstrate successful commissioning.

As a result of the piping rework and water hammer issues at GBF, the BWMS was returned to the BWMS manufacturer following the completion of BE testing. The DP sensor was replaced, and minor software updates were applied. The software updates ensured that internal parameters specific to the installation setup were not lost and ensured the internal clock was not reset to a default time when the system was powered down for a long period, as had occurred at GBF. The manufacturer also optimized the butterfly valve's open and close actuation speed (to avoid water hammers). Because the speed of the butterfly valves was slower after the actuator was adjusted, the start of the backflush operation needed a longer delay to allow the outlet butterfly valve to close completely. The BWMS manufacturer added a longer delay; this minor change to the backwash timing did not change the operating characteristics of the filtration system. Finally, the manufacturer examined the modifications made by GBF to re-route the pipe leaving the BWMS container, which had required the discharge piping to be hung from the ceiling of the container; the manufacturer determined no additional reinforcement was needed to ensure the hanging piping was properly supported.

Importantly, during the examination and repair of the BWMS by the manufacturer, the operational characteristics of neither the filters nor the UV system were changed. Although returning the BWMS to the manufacturer in the middle of the project was not an ideal situation, it was necessary to address the issues discovered during testing at GBF. None of the changes made were expected to alter BWMS' treatment operations or efficacy.

After repairs were made by the manufacturer, the BWMS was shipped directly to GSI. The commissioning tasks completed by GBF and GSI are described in above in Sections 4.2 "Golden Bear Facility Commissioning" and 5.2 "Great Ships Initiative Commissioning" and will not be repeated here. Instead, the next two sections describe high-level differences between the TFs and the ETV Protocol with respect to commissioning.

6.3.1.1 *Golden Bear Facility Commissioning*

GBF personnel developed a suite of SOPs specific to BWMS inspection, installation, commissioning, stressing cycles (i.e., setting the flow rate to 250 m3 h^{-1} for 4 h), and shakedown testing that were included in the TQAP (Golden Bear Facility, 2010). A full-scale, practice BE Test, which would have included sample collection and sample processing, was not conducted during commissioning due to schedule constraints. As a result, additional modifications to the SOPs and the BWMS installation were required after official BE testing commenced, as several unanticipated issues were encountered, including failure of

the differential pressure sensor, more frequent backflushing than anticipated, air bleed into the sample lines, and severe water hammer. This last issue led to manual control of valves at each backflush to reduce water hammer, and ultimately resulted in a second visit from the BWMS technical representative after the second BE test. Following the commissioning tests, both the BWMS technical representative on-site and GBF agreed (with both parties signing the GBF Data [Commissioning] Sheets) that the system was installed properly and ready for operational testing. GBF operated the BWMS without involvement from the vendor's technical representative, reviewed the operation manual, and determined it was sufficiently detailed so that GBF could consistently operate the BWMS. The same was true at GSI.

6.3.1.2 Great Ships Initiative Commissioning

Similar to GBF, GSI developed a suite of SOPs for commissioning the BWMS system at their facility, although GSI did not include the stressing cycles used at GBF. The SOPs documented inspection, installation, checkout, and operational tasks that were conducted to ensure the BWMS was ready for BE Tests. GSI completed a successful commissioning test, including a full-scale test that met the GSI TQAP challenge conditions. Uptake and discharge operations were conducted at a ballast water flow rate of 200 $m^3 h^{-1}$ to the BWMS. GSI used the uptake operation, with augmentation of organisms ≥ 10 µm and < 50 µm, MM, and POC, as an opportunity to collect and analyze biological and water quality samples upstream and downstream of the BWMS. After completing successful commissioning, the technical representative and GSI agreed (with both parties signing the BWMS Commissioning Acceptance Form) that the system was installed properly and was ready for operational testing.

6.3.2 Biological and Water Quality Methods

Although methods for biological and water quality analysis are recommended in the ETV Protocol, TFs may use alternative methods if they are validated and approved by the VO. Table 16 and Table 17 list the methods recommended in the ETV Protocol and those used by GBF and GSI; if an alternate method was used by a TF, the table lists the document validating its use. Finally, if additional methods were used to measure a given parameter, they are listed as well. Note that the treated water sampled by both TFs in the Intercomparison Project contained more organisms ≥ 50 µm than expected, so the methods used by the TFs to enumerate treated samples in this size class were the same as those used for uptake and control tank discharge water. For simplicity, Table 16 lists only the method for uptake and control tank discharge samples for organisms ≥ 50 µm.

Table 16. Biological methods described in the Environmental Technology Verification Protocol and methods used by the test facilities.

	Organisms ≥ 50 µm	**Organisms ≥ 10 µm and < 50 µm**	**Organisms < 10 µm (HPC)**	**Organisms < 10 µm (*Escherichia coli*, Enterococci, and *Vibrio cholerae*)**
ETV Method	Filter samples using sieve with mesh ≤ 50 µm on the diagonal, process immediately by observing with dissecting microscope and probe organisms to determine dead status, fix for total counts and determine counts of living organisms by subtraction	Filter samples with a sieve with mesh ≤ 10 µm on the diagonal, process immediately by epifluorescent staining using two vital, fluorescent stains and manual counts	Dilute samples and process immediately by spreading samples onto two types of recommended media, and incubating them at 25 °C	Process immediately *E. coli*: USEPA Method 1603 or IDEXX Colilert® kit Enterococci: modified USEPA Method 1106.1 or IDEXX Enterolert® kit *Vibrio cholerae*: DNA colony blot hybridization (colonies grown on selective agar and enumerated with a fluorescent antibody kit)
GBF[A] Method	Mesh was acceptable, samples were processed immediately, **living and dead counts were determined simultaneously**	**Mesh was 14 µm on the diagonal, s**taining was the same as ETV, **some samples were analyzed hours later**	**Samples were not diluted**, **1 ETV media was used**, **samples were incubated at room temperature**, samples processed immediately	Same as ETV for *E. coli* and Enterococci (IDEXX Enterolert ® kits) **For *V. cholerae*, colonies were not grown on selective agar,** but the recommended antibody kit was used All were processed immediately
Validation	NR	NR	NR	NR
Alternate Method	ATP (for all size classes)	Staining/flow cytometry ATP (for all size classes)	ATP (for all size classes)	ATP (for all size classes)
GSI Method	Same as ETV	Used one, not two vital stains	Samples diluted for HPC plates, **1 media used**, samples processed immediately	Same as ETV (IDEXX Colilert® and Enterolert® kits)
Validation	NR	Reavie et al., 2010	None - commercial method	NR
Alternate Method	NA	NA	IDEXX SimPlates®	NA

[A]Instances where ETV requirements were not met are shown in bold font. ATP = adenosine triphosphate, ETV= Environmental Technology Verification program, GBF = Golden Bear Facility, GSI = Great Ships Initiative, HPC = heterotrophic plate counts, NA = not applicable, NR = no requirement, TF = test facility, and USEPA = U.S. Environmental Protection Agency.

For the biological analyses, some alternatives to the recommended methods in the ETV Protocol were used (Table 16). In those cases, the VO and AO had concurred with the proposed difference, either for pragmatic reasons or because the change had been validated by the TF. For example, GBF did not dilute samples for heterotrophic plate counts of culturable, aerobic, heterotrophic bacteria (HPC) of organisms < 10 µm prior to spreading the samples on plates. Because their experience showed counts were naturally low and did not require dilution, the VO concurred that not diluting the samples was acceptable for this research project. Likewise, it was agreed GBF could forego using two agars for the HPC for organisms < 10 µm or growing *Vibrio cholerae* colonies on agar before using fluorescent antibodies to detect *V. cholerae* because GBF did not have the capacity to implement either procedure before testing commenced. GBF did later provide a memo from the manufacturer of the *V. cholerae* test kits citing their low detection limit (1 bacterium 100 mL^{-1}) without growing colonies on agar. The memo, however, also stated that the samples GBF sent to the manufacturer for analysis had been stored for 6-8 weeks, and the results indicated the presence or absence of the bacteria at the time of analysis, not the time of collection.

For heterotrophic plate counts (HPC), GBF incubated the samples at room temperature, as they did not have the proper incubator to control the temperature; GBF plans to use an incubator in future testing. Importantly, these conditions render the HPC data suspect. For the concentration of organisms ≥ 10 µm and < 50 µm, GBF's oversight of using slightly larger mesh (10 µm) than specified in the ETV likely had a small effect on the data, as the protist community in Carquinez Strait is dominated by larger protists. In subsequent testing, GBF plans to use 7 µm mesh (10 µm in diagonal).

One alternative to the recommended ETV biological methods was used by GSI. For one of the agars used in HPC analyses, a patented, commercial test kit (IDEXX SimPlate®) was substituted. The VO concurred with this change.

It should be noted that the ETV Protocol does not specifically address the use of positive and negative controls for the microbiological samples (i.e., all analyses for organisms < 10 µm). This oversight should be corrected, as such controls are necessary in any microbiological analyses to ensure spurious results do not occur due to contamination, expired materials, or other factors. During the AO's pre-test audit of GSI and monitoring of BE tests at both facilities, however, the need for positive and negative controls was discussed. Although GBF staff and facilities could not accommodate an audit in advance of testing, in their feedback following testing, the AO pointed out to GBF the need for positive and negative controls for all microbiological tests (e.g., the HPC analyses employed no controls), the need for training records for all Analysts (which were not available), and the need to provide calibration records for all instruments, laboratory pipettes, and glassware (which also were not available). Following the audit at GSI and after monitoring the BE tests, the AO also provided suggestions for relatively minor improvements. More detail can be found in the report to USCG-RDC by NSF (NSF, in preparation).

Regarding the water-quality measurements in Table 17, the VO concurred with both TF's use of calibrated, commercial sondes to measure dissolved oxygen. Samples collected for water quality analyses by GBF were not stored according to EPA methods (e.g., frozen or refrigerated). Without a study to show that samples did not degrade under the conditions, the data should be considered questionable.

Table 17. Water quality methods described in the Environmental Technology Verification Protocol and methods used by the test facilities.

Protocol or TF	Method for dissolved organic carbon	Method for particulate organic carbon	Method for total suspended solids	Method for dissolved oxygen
ETV	Filter sample through a GF/F, freeze filtrate up to 28 d, analyze using a carbon analyzer and an APHA, ASTM, or EPA method listed in ETV Protocol	Filter sample through a GF/F, freeze filter up to 28 d, analyze using a carbon or CHN analyzer and an APHA or EPA method listed in ETV Protocol	Analyze sample immediately or refrigerate up to 7 d, analyze using a 5-place balance and an APHA or EPA method listed in ETV Protocol	Collect sample, fix, and titrate within 24 h using a Radiometer TitraLab® and APHA or EPA method listed in ETV Protocol
GBF[A]	Filtered through a GF/F into a glass DOC vial containing an **acid preservative; refrigerated** until delivery to an EPA-approved laboratory on the same day as sample collection; analyzed the next day using a total organic carbon analyzer following an EPA standard method[B]	Filtered sample through a GF/F and dried immediately in 65°C oven in marked plastic envelopes for 48 h; held in a vacuum desiccator at **room temperature** until analysis on a CHN analyzer using a standard EPA method	**Analyzed filter after holding in a desiccator at room temperature up to 30 d** using a 5-place balance and an EPA standard method[B]	**YSI Sonde[C] and in-line sensor** was identified by TQAP, but according to SOP redlines was not used, and grab samples were not analyzed. Only source of data was continuous inline sensor (Sea-Bird Thermosalinograph)
GSI	Filtered through a GF/F, analyzed on the day of sample collection using a Total Organic Carbon Analyzer and an APHA method	POC was determined empirically by first measuring the NPOC and DOC concentrations, then performing the calculation: NPOC – DOC = POC	Analyzed refrigerated on the day of sample collection using a 5-place balance and APHA method	**YSI Sonde and in-line sensor**, (in-line sensor measurements were not reported, as they were deemed by GSI to be inaccurate)

[A]Instances where ETV requirements were not met are shown in bold font. [B]The SOP for this measurement was not followed; in post-test discussions, the method in this table was described to the VO. [C]NRL concurred with both TFs' use of calibrated, commercial sondes or sensors to measure dissolved oxygen. APHA = American Public Health Association, ASTM = ASTM International, CHN = carbon/ hydrogen/nitrogen, DOC = dissolved organic carbon, EPA = Environmental Protection Agency, ETV = Environmental Technology Verification program, GBF = Golden Bear Facility, GF/F = glass fiber filter, GSI = Great Ships Initiative, NPOC = non-purgeable organic carbon, and POC = particulate organic carbon.

6.3.3 Results

6.3.3.1 *Statistical Analyses*

To determine if the BE tests results showed statistically significant differences in the estimates of organism abundances in intake and discharge samples among BE tests within a facility or between facilities, data were analyzed using Minitab 16 Statistical Software (2010, Minitab, Inc., State College PA). Because all sample types (uptake, control discharge, and treatment discharge) contained non-sparse communities, the data analysis was first attempted using parametric tests (i.e., tests that assume a normal distribution of samples). Initial normality tests indicated that the data were a mix of non-normal and normal samples. This result was partly due to the very small ($n < 5$) sample sizes for some of the analyses. Indeed, very small sample sizes substantially reduce the power and efficacy of normality testing and may fail to identify data as non-normal (i.e., reject the null hypothesis of normality) even if the data are non-normal (Whitlock and Schluter, 2009). For the BE test data, subsequent manipulation revealed the data could not be transformed to another distribution (distributions attempted included lognormal, logistic, log logistic, exponential, gamma, Weibull, as well as the Box-Cox transformation). Thus, the best course of action was to use nonparametric methods for analysis rather than parametric methods (e.g., analysis of variance, ANOVA). As mentioned above, with very small sample sizes (particularly where $n < 5$), it is nearly impossible to reliably determine if the data follow a normal distribution, and it would be fundamentally indefensible to make statistical comparisons using a test that assumes normality (e.g., ANOVA).

The specific nonparametric methods used were Mann-Whitney U tests (a nonparametric method for hypothesis testing) and Kruskal-Wallis tests (a nonparametric alternative to ANOVA, which uses the median as a measure of central tendency rather than the mean). Given that the Kruskal-Wallis tests involved multiple comparisons, the familywise error rate must be considered. The familywise error rate is defined as the maximum probability that at least one comparison will falsely conclude that at least one of the observed differences is significantly different from the null hypothesis. Restated, with multiple pairwise comparisons, the chances for committing a type I error (i.e., a false positive) are higher than the error rate for any individual comparison, and as a result, the individual alpha level for tests must be adjusted accordingly to maintain an overall family error rate. Given the small sample sizes present in the data, a familywise error rate of 0.20 was selected, which required individual alpha levels for tests ranging from 0.013 to 0.033 (the typical alpha level used in scientific analyses is 0.05). For comparative purposes, analyses were run with a familywise error rate of 0.05; these results, which yielded fewer statistically different comparisons than when 0.20 was used, are included in Appendix C to this report).

Nonparametric methods do not assume the data follow a distribution, and particularly with small sample sizes, they are less powerful than their parametric alternatives (Johnson and Bhattacharyya, 2006). That is, they are potentially unable to detect a significant difference when one is present (a type II error, i.e., a false negative). As a result, it is not surprising that the nonparametric tests did not detect significant differences among some BE tests even though the differences appeared to be substantial (e.g., when viewing a histogram of the data means, as in Figure 12). If there were more observations per sample (i.e., if n was higher), it is possible the tests would yield more powerful information (detect more significant differences) or, potentially, allow for parametric methods, such as the relatively robust ANOVA, to be used (Whitlock and Schluter, 2009).

It is recommended that facilities conduct a prospective power analysis to estimate the appropriate sample size before carrying out analyses of uptake and control samples, which are expected to employ statistical tests that assume a normal distribution of the data (it is assumed that treated samples would contain sparse populations and thus adhere to a Poisson distribution, and a power analysis would be unnecessary). Power analysis requires, first, knowledge of the "effect size" to be examined, which is defined as the minimum detection difference between groups, e.g., is it necessary to know the concentration of living organisms ≥ 50 µm is 100,000 m^{-3} vs. 100,010 m^{-3}? Alternatively, perhaps detecting a larger difference, say between 100,000 m^{-3} and 90,000 m^{-3}, is sufficient. Second, power analysis requires an estimate (which can potentially be determined from historical data) of the standard deviation around the mean of the measurement (Lenth, 2001). Both parameters are used to compute the power function and estimate sample size. It is a statistical convention to design experiments with a power of 0.80 (Whitlock and Schluter, 2009). As such, there is a 20% chance the experiments will commit a type II error (the analysis is unable to detect a significant difference when one actually exists). Notably, while larger sample sizes are vastly preferable, the practicality of conducting a prospective power analysis (particularly determining the effect size) or substantially increasing sample sizes for all samples might be cost prohibitive. In the event that a prospective power analysis cannot be performed to determine necessary sample sizes, from the data collected in this study, it is recommended that each sample of organisms from the two largest size classes contain at least five measurements (i.e., ≥ 5 subsamples) to conduct robust and informative normality and hypothesis tests on the data.

It is noteworthy that in enumerating organisms ≥ 10 µm and < 50 µm at GSI, one sample was analyzed per BE test. Hence, sample sizes were one, which creates an environment where it is impossible to perform nonparametric analyses. Consequently, GBF data from each of the three BE tests are compared to aggregated GSI data from all three BE tests (essentially, n = 1, with 3 subsamples).

6.3.3.2 *Biological Data*

Using the above approach, nine comparisons were made between facilities: three water types (uptake, control discharge, and treatment discharge) for each of the three size classes (living organisms ≥ 50 µm; living organisms ≥ 10 µm and < 50 µm determined using epifluorescent stains; and living organisms < 10 µm). The nine comparisons are graphed below (Figures 12-20) with a familywise error rate of 0.20; note that the units on the y-axes vary among graphs. Although the statistical analysis distinguished differences among ranked data, for clarity, the mean values were graphed. Groups that are statistically different are indicated by different letters (a, b, and c) above the bars, e.g., in Figure 12, GBF BE Test 1 is significantly different from the other two GBF tests, but it is not significantly different from any of the GSI BE tests. Also, where appropriate, the dashed line indicates the requirement from the ETV Protocol, and the solid line represents the DFS for the facility. For brevity, the entire statistical output of these tests, including p-values, and tests performed with a familywise error rate of 0.05, is not included in the main body of this report but can be found in Appendix C.

The data presented for organisms < 10 µm are the HPC from GBF and the IDEXX SimPlate® results from GSI (Figures 18-20). Ideally, the plate counts from both facilities would be compared, but at GSI, data contamination of the HPC blanks during the third BE test at GSI rendered that data questionable, so the SimPlate® data are graphed below; the HPC data from GBF are also questionable because the incubation temperatures were out of specification (see Section 6.4.2 "NSF International Audits"). Although the HPC and SimPlate® data from GSI varied approximately 2-5 fold for a given test (Table 5), the mean concentration of organisms < 10 µm was generally orders of magnitude greater at GSI than GBF (Figures 18-20).

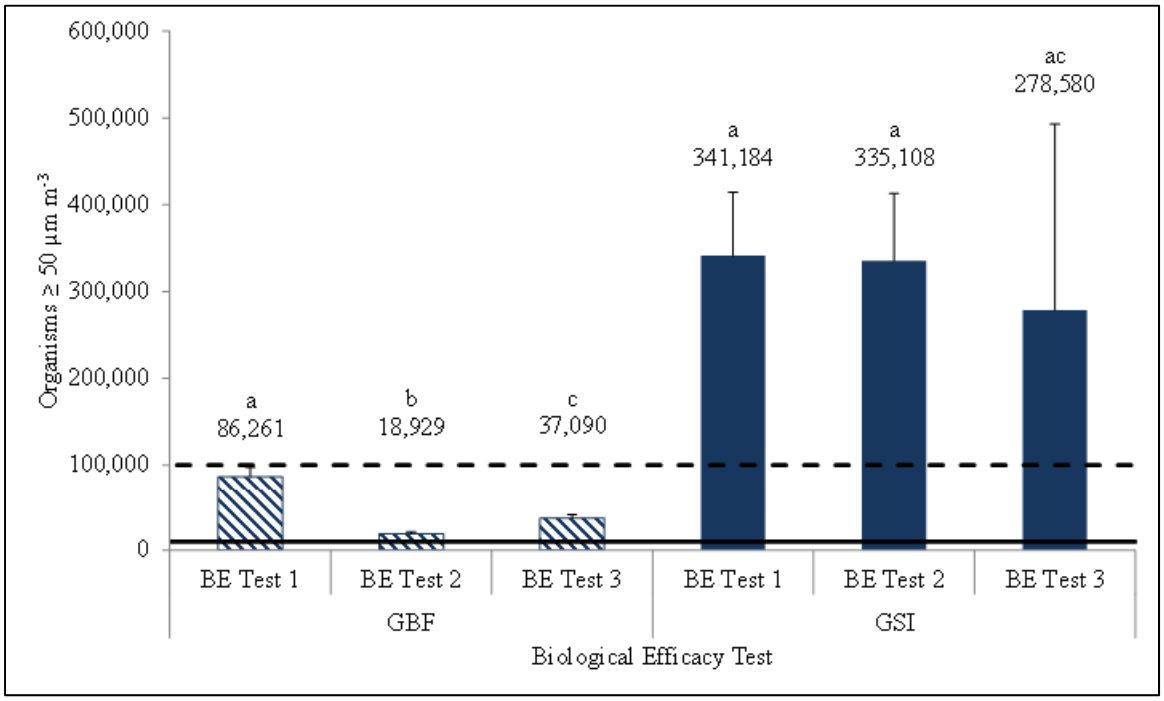

Figure 12. Living organisms ≥ 50 µm in uptake water. Statistically different comparisons are indicated by different letters; numbers above the bars represent their value; error bars represent one standard deviation; the dashed line indicates the requirement of 100,000 m^{-3} in the Environmental Technology Verification program Protocol, and the solid line represents the deviation from specification to GBF of 10,000 m^{-3}. BE = biological efficacy, GBF = Golden Bear Facility, and GSI = Great Ships Initiative.

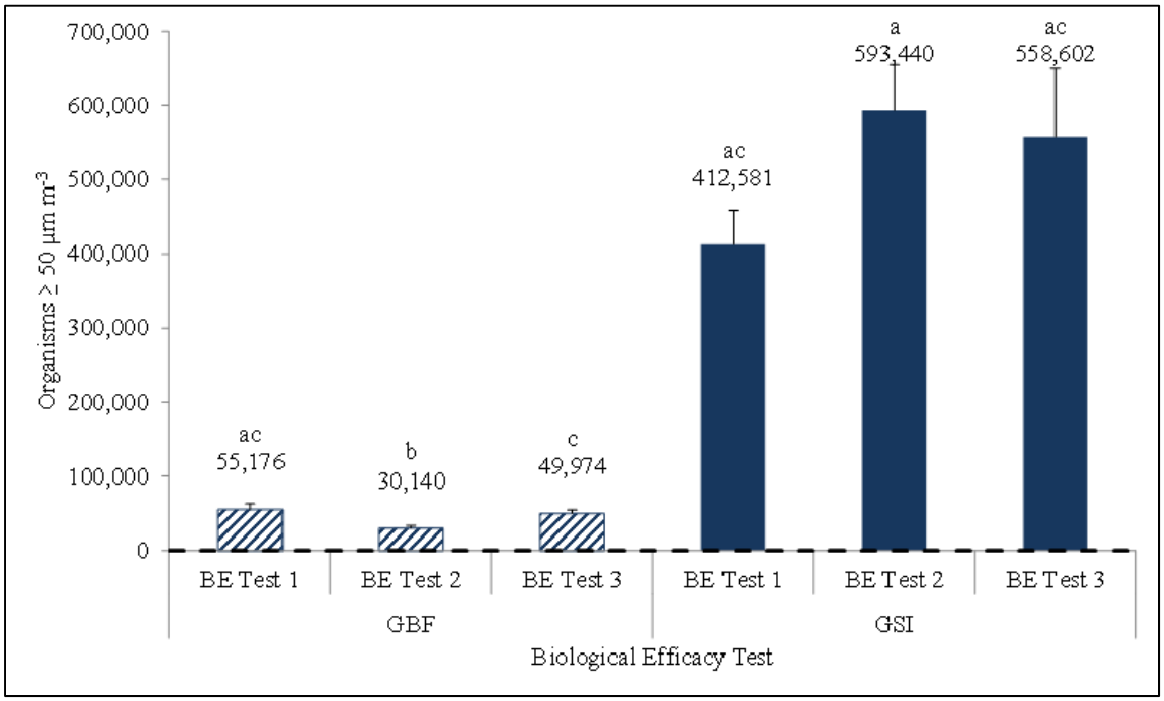

Figure 13. Living organisms ≥ 50 µm in the control tank discharge. The dashed line indicates the requirement in the ETV Protocol of 100 living organisms m^{-3}.

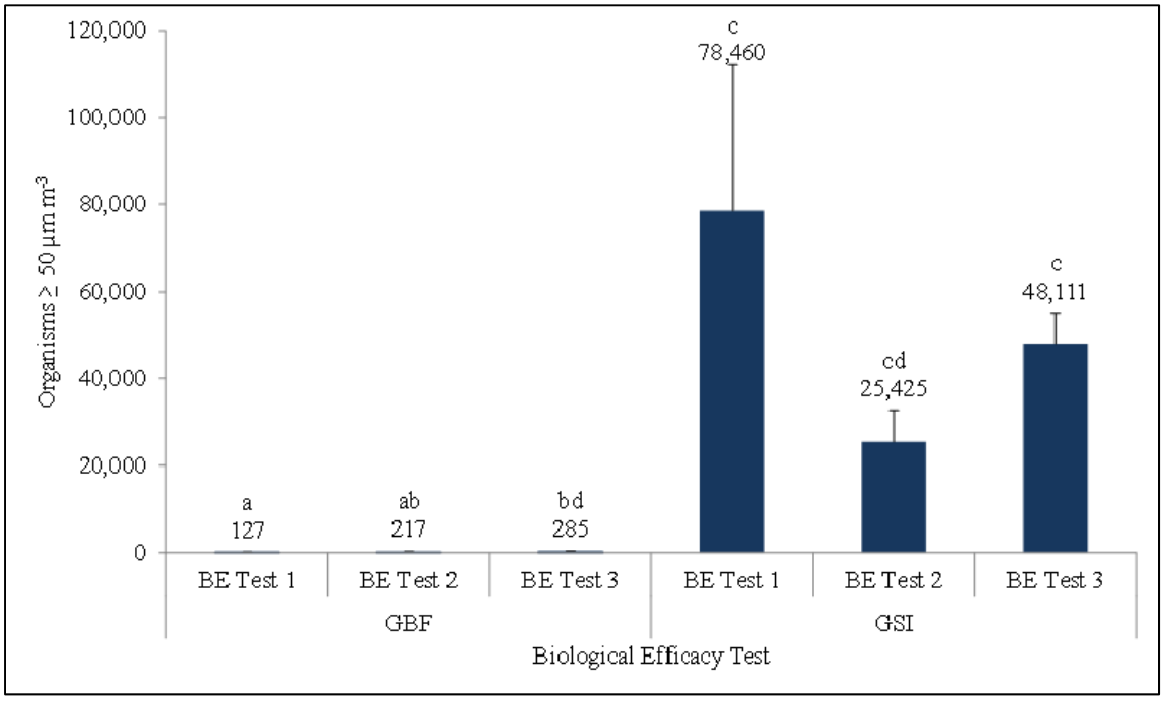

Figure 14. Living organisms ≥ 50 µm in the treatment tank discharge.

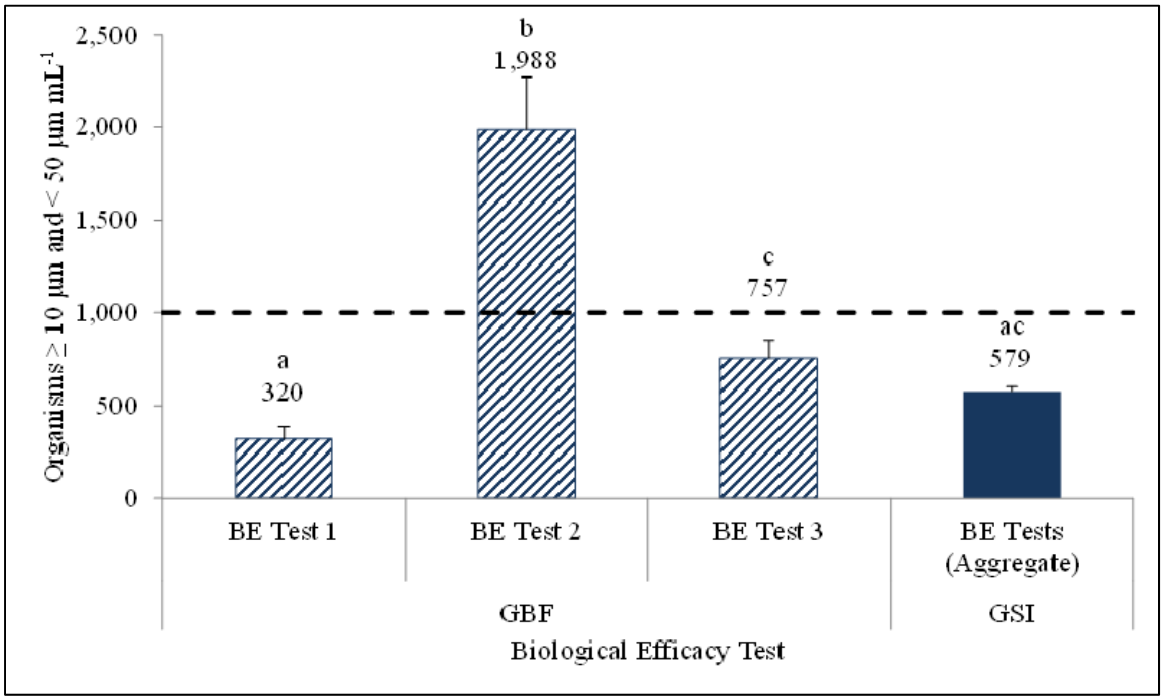

Figure 15. Living organisms ≥ 10 µm and < 50 µm in uptake water quantified using epifluorescence microscopy. The dashed line indicates the requirement in the ETV Protocol of 1,000 mL^{-1}.

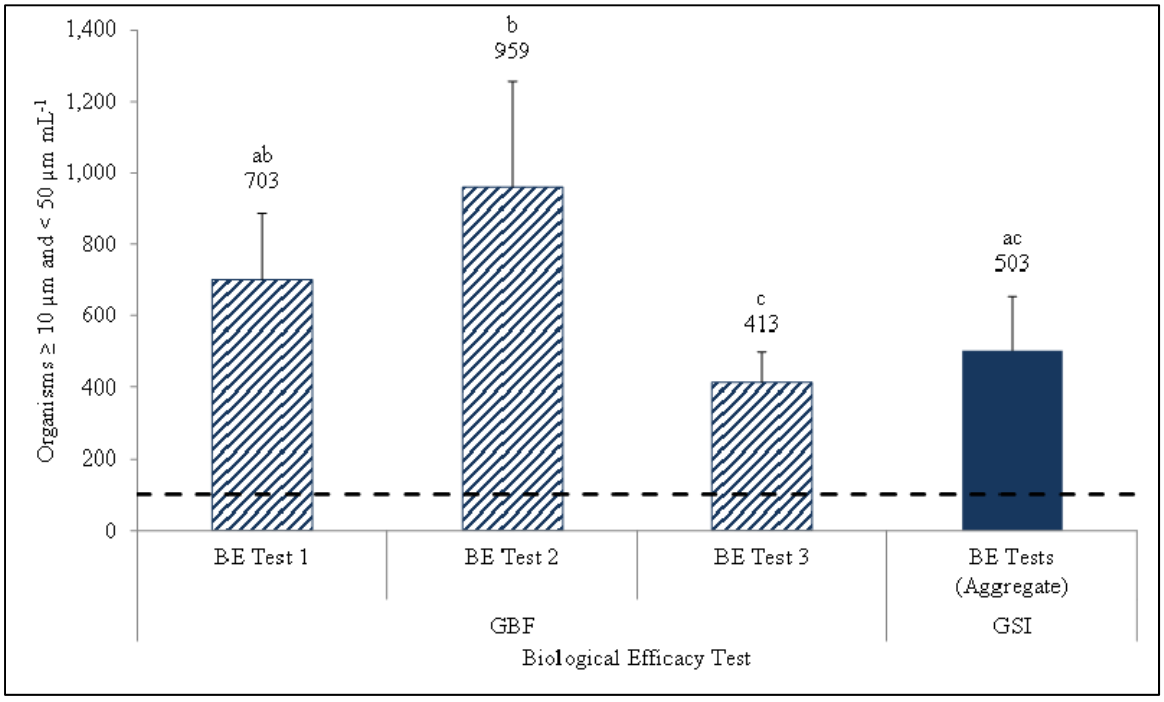

Figure 16. Living organisms ≥ 10 µm and < 50 µm in the control tank discharge quantified using epifluorescence microscopy. The dashed line indicates the requirement in the ETV Protocol of 100 mL^{-1}.

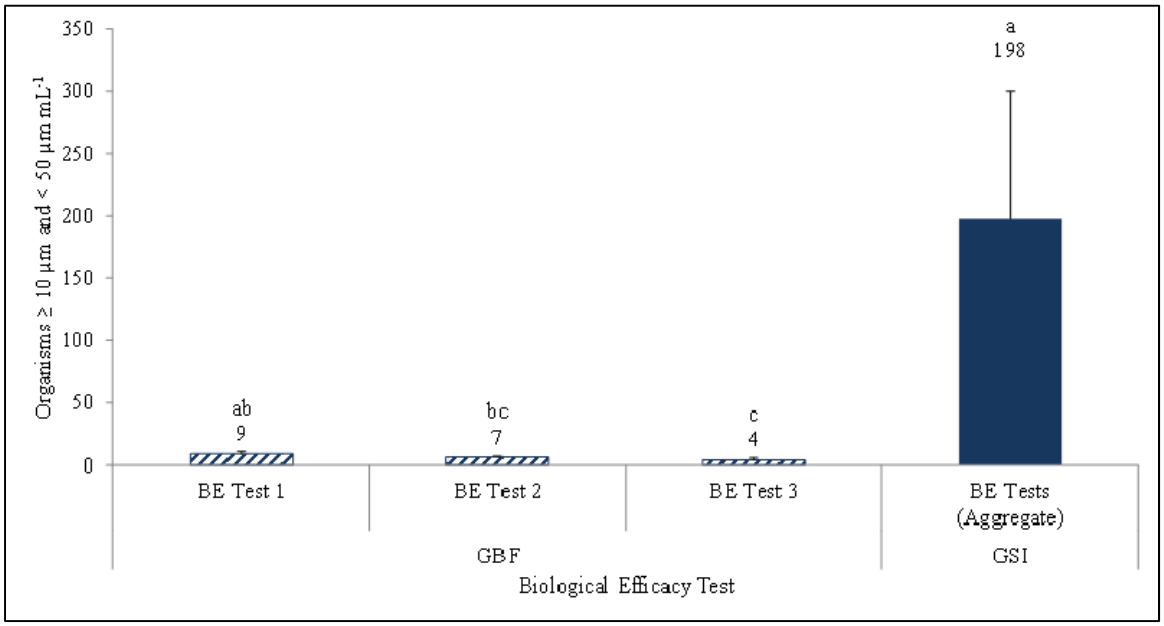

Figure 17. Living organisms ≥ 10 µm and < 50 µm in the treatment tank discharge quantified using epifluorescence microscopy.

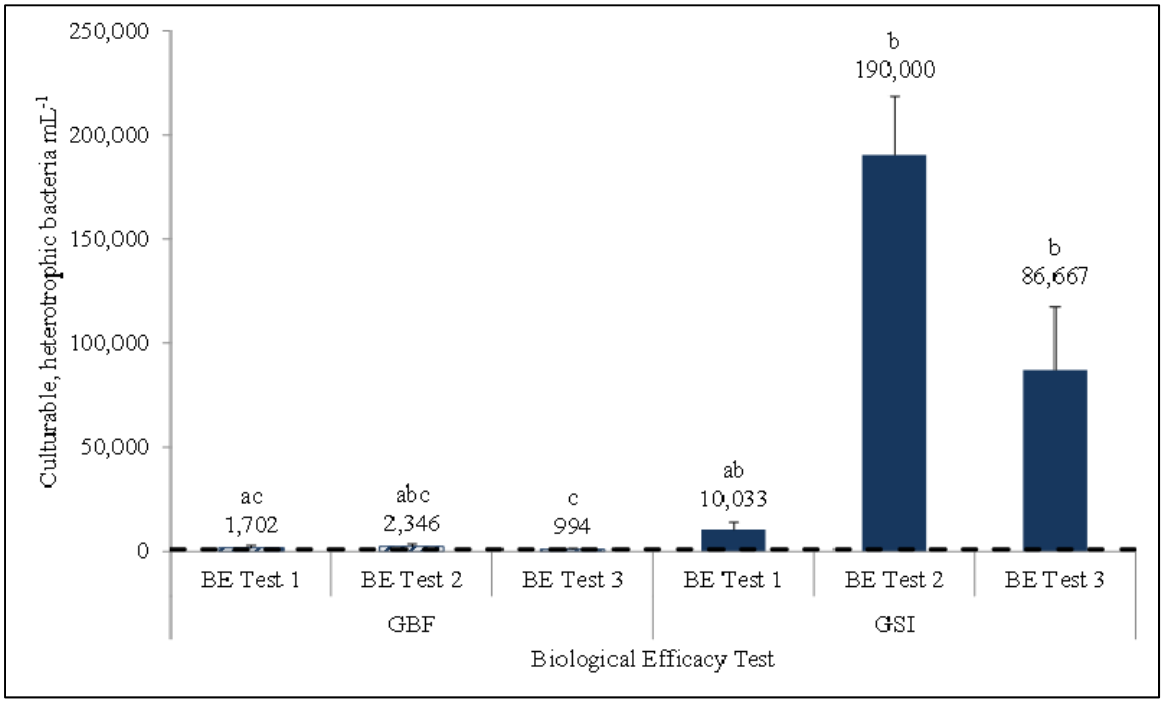

Figure 18. Living organisms < 10 µm in the uptake water. The dashed line indicates the requirement in the ETV Protocol of 1,000 mL^{-1}.

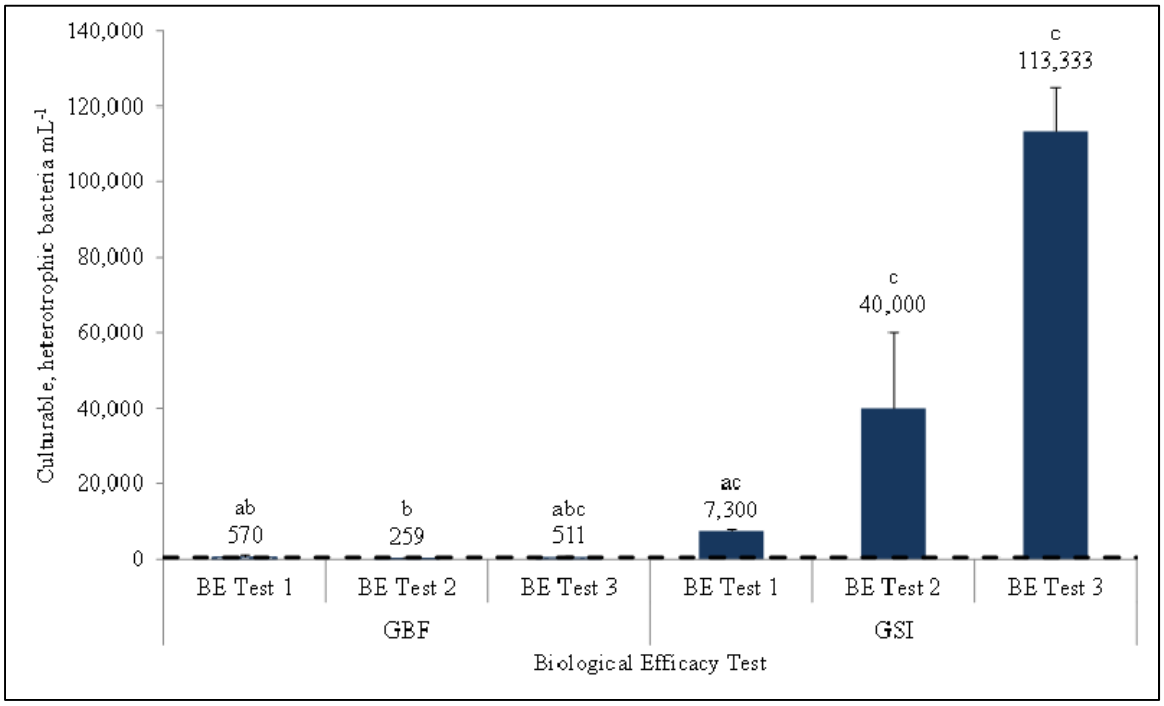

Figure 19. Living organisms < 10 µm in the control discharge. The dashed line indicates the requirement in the ETV Protocol of 500 mL^{-1}.

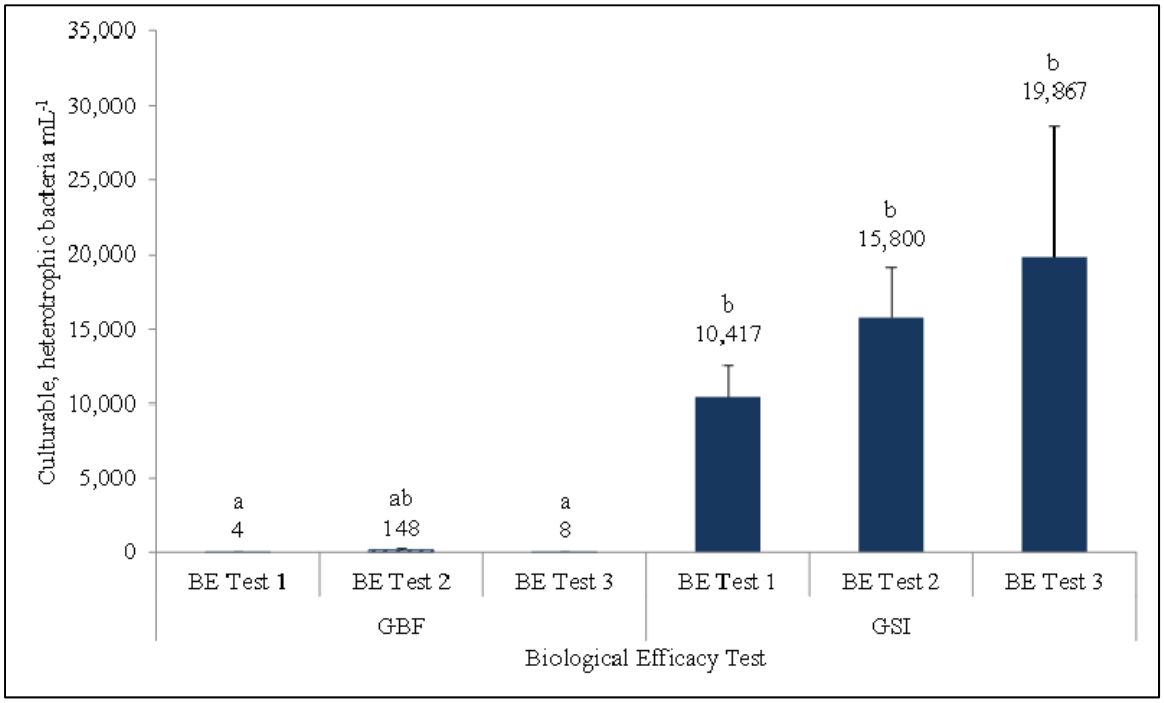

Figure 20. Living organisms < 10 µm in the treatment discharge water.

The ETV Protocol lists minimum requirements for the concentration of living organisms in the control tank discharge. At GBF, the requirements were met in all three tests for the two largest size classes, although in BE tests 1 and 2, the mean concentration of organisms < 10 µm was below the requirement of 1,000 mL^{-1} (Figures 13, 16, and 19, and Table 3). In all three BE tests, GSI met the control tank discharge requirements (Figures 13, 16, and 19, and Table 5).

The Kruskal-Wallis tests showed that for seven out of the nine comparisons, there were statistically significant differences between GBF and GSI, and the concentrations of organisms reported by GSI tended to be larger than those reported by GBF. The exceptions to this general rule were found in the uptake and control tank discharge results for organisms ≥ 10 µm and < 50 µm (Figure 15 and Figure 16), where the data tended to not be significantly different between TFs. These findings are supported by Mann-Whitney U tests, which compared the aggregate GBF data to the aggregate GSI data and detected significant differences in seven of nine groups, with once again the ≥ 10 µm and < 50 µm uptake and control tank discharge tests not displaying significant differences ($\alpha = 0.05$). In none of the nine comparisons were all BE tests from one TF significantly different from all the BE tests at the other TF (e.g., all of the data bars in a plot from GBF had 'a's above them, and all the data bars from the GSI had 'b's above them).

One potential issue for future analysis is the overall representativeness of the sample. For the purposes of this report, it is relatively unimportant due to the high concentrations of organisms in all samples (control and discharge); however, for sparse populations of organisms in treated water, ballast water residuals within the tank can be a concern. If a disproportionately high or low concentration of organisms is found in the remaining water in the tank as compared to the water sampled upon discharging a tank, the representativeness of the sample would be called into question. Without a representative sample, the statistical methodology applied to sampling is potentially invalidated. As a result, further research into this issue is recommended, and TFs should strive to minimize residual volumes of ballasted water in tanks.

The data for organisms ≥ 50 µm at GSI show an interesting feature: the mean numbers of organisms in the control discharge water are higher than the means in the uptake water (Figure 13 and Figure 12). Given that the community was dominated by rotifers, this result suggests rotifer reproduction (by eggs hatching during the 48-h hold time) exceeded rotifer mortality, or some rotifers < 50 µm grew during the hold time so they were ≥ 50 µm upon discharge. This aspect of the biota, likely typical in freshwater systems, illustrates the liability of a BWMS that treats water near a given discharge standard.

Another way to compare the results between facilities is to examine the change in the density of organisms in the uptake water vs. the treatment discharge water (Table 18). Here, the data are shown as log (10-fold) reductions from uptake to discharge following treatment. First, the $\log_{10}$ value was calculated for each average concentration of organisms for each size class. The log reduction was then determined by subtracting the uptake $\log_{10}$ concentrations from the treatment discharge $1\log_{10}$.

Table 18. Log change in the concentration of living organisms between uptake and treatment discharge at each facility.

Living organisms by size class (method, where applicable)	GBF log change between uptake and treatment tank discharge	GSI log change between uptake and treatment tank discharge
≥ 50 µm	BE 1: -2.8 BE 2: -1.9 BE 3: -2.1	BE 1: -0.6 BE 2: -1.1 BE 3: -0.9
≥ 10 and < 50 µm (epifluorescent microscopy)	BE 1: -1.5 BE 2: -2.5 BE 3: -2.2	BE 1: -0.3 BE 2: -0.5 BE 3: -0.8
< 10 µm (HPC for GBF, IDEXX® plates for GSI)	BE 1: -2.6 BE 2: -1.2 BE 3: -2.1	BE 1: 0.0 BE 2: -1.1 BE 3: -1.6

BE = biological efficacy test, HPC = heterotrophic plate counts, GBF = Golden Bear Facility, and GSI = Great Ships Initiative.

With the exception of organisms < 10 µm in BE Test 1 at GSI, all values were negative, indicating a decrease in the concentration of organisms following treatment. Furthermore, the magnitude of the reduction was relatively consistent within a TF (GBF's results varied from -1.2 log to -2.8 log, and GSI's results varied from 0.0 log to -1.6 log). Thus, this treatment effect was evident at both facilities, despite differences in the biological community in uptake water, both with respect to the number and composition of organisms. That said, the facilities' results differed in that in all cases, and the magnitude of the reduction was greater at GBF than at GSI.

6.3.3.2.1 Effect of transmittance

Percent transmittance (or transmissivity) is an important parameter for BWMSs that incorporate UV treatment, as decreasing optical transmission requires increasing the energy output from the UV source to maintain UV dosage. Both TFs incorporated some means to measure percent transmittance during BE testing. GBF used an in-line sensor to continuously monitor uptake water transmissivity. In the GBF verification report, there was no reference to these data or comparison of transmissivity data to other treatment performance results, but plots were provided in Appendix H that charted approximately 2 h and 45 min of transmissivity data for each of the three uptake cycles. No indication of beginning and end of uptake was provided (the uptake duration was 2 h 15 min for BE test 1 per Table 3-1 in the GBF verification report [Golden Bear Facility, 2012]). Of note, the plots for BE test 2 uptake (18 JAN 2011) and BE test 3 uptake (25 JAN 2011) were identical; the data for 18 JAN 2011 are likely incorrect based on visual comparisons to corresponding fluorometry data collected over the same time period. From these plots, the uptake water transmittance data for BE test 1 was less than 6.5% for the duration of the test, and the transmittance for BE test 3 ranged between 12.5% and 24.5% over the duration of the test. Because the data logs were not recorded properly by the BWMS, the UV lamp intensity cannot be correlated with the transmittance values; however, if the BWMS was working properly—and spot checks on site showed that to be true—the UV intensity would have automatically adjusted to maximum dosage at low transmittance values.

GSI measured percent transmission of multiple grab samples collected during uptake and treatment discharge; both filtered and unfiltered samples were analyzed using a spectrophotometer. Results from each test were not provided in the Executive Summary Validation Matrix; however, individual sample data were presented in the Results section. The data for transmittance were similar for control, pretreatment, and post-treatment samples and similar between samples that were unfiltered or filtered (with a glass fiber filter, nominal pore size 0.7 µm). Typical results for BE test 1 were 15.9% filtered, 14.2% unfiltered; for BE test 2 were 27.5% filtered, 24.2% unfiltered; and BE test 3 were 14.8% filtered and 10.9% unfiltered. Where measured, treatment discharge transmissivities were within 1% of the uptake values for both filtered and unfiltered samples. The potential effect of local challenge conditions for percent transmission as compared to ETV requirements was discussed by GSI in the verification report Section 9.2, "GSI Facility Operation and Protocol Implementation" (Great Ships Initiative, 2012).

6.3.4 BE Validity Criteria

Each TF presented a Validity Matrix in their TQAP to indicate the valid ranges of operational, biological and water chemistry parameters for a BE test cycle. These values included the DFSs to challenge conditions previously described (Table 2). Subsequent to each BE test, each TF populated the validity matrix with test data to show whether the requisite challenge and operational conditions were met. Completed matrices were to be provided at the end of each test with preliminary data to provide an initial assessment of the test's validity.

Minimum requirements for validity matrix parameters are provided in the Protocol. These parameters include flow, pressure, and volume ranges for both ballasting and sampling operations; ranges for water chemistry core parameters; and densities for living organisms in each size class. Each TF was free to specify additional parameters to define a valid test cycle. The validity matrices created by GBF and GSI are shown in Figure 21 and Figure 22, respectively.

Test Cycle Number:

PIC _____ Print _____ Sign _____ Date

Quality _____ Print _____ Sign _____ Date

SOP 1, Step 3 Validating Test Parameters

- Validate Test Parameters in below log:
  - Uptake _____.
  - Treatment Discharge _____.
  - Control Discharge_____.

	Criteria	Uptake Cycle	Discharge Cycles Treatment	Discharge Cycles Control
Treatment Line and Tank				
Average (m3/hr) (Note 1)	200 +/- 10%			
Volume at end Cycle (m^3)	400 +/- 10%			
Control Line and Tank				
Total Volume (m^3)	400 +/- 10%			
Combined Sample Volume (m^3)	(Note 2)			
Uptake and Control Discharge	≥ 3			
Treatment Discharge	≥ 9			
Ballast Hold Duration (hours)	48 +/-10%			
Water Quality				
Salinity (PSU)	10 - 20			
Temperature (Celcuis)	4 - 35			
DOM (mg/L)	≥ 3			
POM (mg/L)	≥ 2			
TSS (POM + MM) (mg/L)	≥ 20			
Uptake Living Populations				
≥50 microns (organisms/m^3)	10^4			
≥10 < 50 microns (organisms/mL)	10^2			
<10 microns (bacteria/mL)	10^3			
Control Living Populations				
≥50 microns (organisms/m^3)	100			
≥10 < 50 microns (organisms/mL)	100			
<10 microns (bacteria/mL)	500			

Ballasting Operations			
(Live/m^3)	*Dominant Species #3*		
	Dominant Species #4		
	Dominant Species #5		
	Total Live Density		
≥10 to <50µm Diversity and Concentration (Live/mL)	Valid Range	5 species across 3 phyla and ≥1,000/mL (Total Live Density)	
	Dominant Species #1		
	Dominant Species #2		
	Dominant Species #3		
	Dominant Species #4		
	Dominant Species #5		
	Total Live Density		
<10 µm Concentration (Heterotrophic Bacteria/mL)	Valid Range	≥1000 MPN or CFU/mL	
	Mean Measured Value		

Were all measured values during Ballasting Operations within the acceptable valid ranges? ☐ **YES** ☐ **NO**

COMMENTS:

Testing Organization Representative Signature/Date:

X _____

Kelsey R. Prihoda
GSI Quality Assurance/Quality Control Officer

Figure 21. Validation matrices from the Golden Bear Facility.

Great Ships Initiative

GSI/LB/QAQC/TQAP/GFE
Revision No. 1: July 22, 2011
Page 763 of 765

Appendix 9: Great Ships Initiative (GSI) Biological Efficacy (BE) Testing Validation Matrix

Test Cycle ID Code:

Intake (Ballasting) Operations

Date/Time of Intake (Ballasting):

Ballasting Operations

Parameter		Control Tub	Pre-Treatment Tub	Control Retention Tank	Treatment Retention Tank
Circle Sampling Location:		1-SP2A 2-SP2B 3-SP2C	4-SP3A 5-SP3B	C1 C2	T1 T2
Flow Rate (m^3/hr)	Valid Range of Average	2.0 to 3.6		140 to 170	180 to 210
	Avg. Measured Value				
Pressure (psi)	Valid Range of Average	35 to 50			
	Avg. Measured Value				
Volume (m^3)	Valid Range	2.6 to 3.6		175 to 205	
	Measured Value				

Water Quality Conditions

Parameter		Control In-Line	Pre-Treatment In-Line
Circle Sampling Location:		SP2A SP2B SP2C	SP3A SP3B SP3C
Temperature (°C)	Valid Range of Average	4 to 35	
	Avg. Measured Value		
Salinity (ppt)	Valid Range of Average	≤1	
	Avg. Measured Value		
pH	Valid Range of Average	6 to 9	
	Avg. Measured Value		
DO (mg/L)	Valid Range of Average	6 to 11	
	Avg. Measured Value		
DOM (mg/L as DOC)	Valid Range of Average	≥6	
	Avg. Measured Value		
POM (mg/L as POC)	Valid Range of Average	≥4	
	Avg. Measured Value		
MM (mg/L)	Valid Range of Average	≥20	
	Avg. Measured Value		
TSS (mg/L)	Valid Range of Average	≥24	
	Avg. Measured Value		

Biological Diversity and Concentrations

Parameter		Control	Pre-Treatment
Circle Sampling Location:		1-SP2A 2-SP2B 3-SP2C	4-SP3A 5-SP3B
≥50 µm Diversity and Concentration	Valid Range	5 species across 3 phyla and ≥100,000/m^3 (Total Live Density)	
	Dominant Species #1		
	Dominant Species #2		

GSI/LB/QAQC/TQAP/GFE
Revision No. 1: July 22, 2011
Page 765 of 765

Discharge (Deballasting) Operations

Date/Time of Discharge (Deballasting):

Deballasting Operations

Parameter		Control Tub	Treatment Tub	Control Retention Tank	Treatment Retention Tank
Circle Sampling Location:		1-SP10C 2-SP10B 3-SP10A	4-SP10C 5-SP10B 6-SP10A	C1 C2	T1 T2
Retention Time (hr)	Valid Range	Not Applicable		43.2 to 52.8	
	Measured Value				
Flow Rate (m^3/hr)	Valid Range of Average	3.3 to 3.6		190 to 210	
	Avg. Measured Value				
Pressure (psi)	Valid Range of Average	30 to 40			
	Avg. Measured Value				
Volume (m^3)	Valid Range	3.3 to 3.6		Not Applicable	
	Measured Value				

Biological Concentrations

Parameter		Control Tub	Treatment Tub
Circle Sampling Location:		1-SP10C 2-SP10B 3-SP10A	4-SP10C 5-SP10B 6-SP10A
≥50 µm Concentration (Live/m^3)	Valid Range	≥100/m^3 (Total Live Density)	Samples Collected? (YES/NO)
	Total Live Density		
≥10 to <50µm Concentration (Live/mL)	Valid Range	≥100/mL (Total Live Density)	Samples Collected? (YES/NO)
	Total Live Density		
<10 µm Concentration (Heterotrophic Bacteria/mL)	Valid Range	≥500 MPN or CFU/mL	Samples Collected? (YES/NO)
	Mean Measured Value		

Were all measured values during Deballasting Operations within the acceptable valid ranges? ☐ YES ☐ NO

COMMENTS:

Testing Organization Representative Signature/Date:

X _____

Kelsey R. Prihoda
GSI Quality Assurance/Quality Control Officer

Figure 22. Validation matrices from the Great Ships Initiative.

Each TF wrote their validation matrix as a form to be completed during testing. GSI provided more criteria and more detailed parameters for validating a test. This approach seems preferable, as it provided a more detailed justification for the validity of a test.

In practice, GBF completed portions of the form shown in Figure 21 on separate pages for the operations (i.e., uptake, control discharge, and treatment discharge), so a single sheet with all data for a given BE test was not available. Preliminary results for each operational cycle were verbally reported to the VO, raw data for operational and summary biology data were later sent to the VO in spreadsheets, and GBF subsequently assembled this information into a validation matrix using the form shown in Figure 21 for all three tests. Conversely, GSI used a single form so that uptake data were reported following the uptake operation, and the discharge data were added to a different page subsequent to the discharge. These pages were provided to the VO following the respective uptake and discharge operations by GSI. Completed validation matrix sheets at both facilities were scanned into their respective databases during testing, and weekly discussions were held with each facility to review preliminary results following each test cycle. The Executive Summary section of the verification report from each facility provides the final version of the verification matrix showing results from all three BE tests for each TF.

In general, both TFs met their respective validity criteria with a few minor exceptions. GBF did not reach some challenge water chemistry criteria and was low on some biology counts; GSI did not meet control flow criteria on two fill cycles, was slightly low in treatment flow on one fill cycle, and collected slightly more discharge sample volume than planned on most runs (Table 3 and Table 5). However, in the VO's opinion, these deviations did not affect the validity of the discharge measurements or the tests at either facility.

6.4 Quality Assurance/Quality Control

6.4.1 Data Quality Indicator Matrix

As part of the QA/QC process, the ETV program requires that the quality of data collected during testing is assessed using statistical analysis. Data quality assessment is unique to each facility, although guidance is provided in the ETV Protocol, Appendix A, "Quality Assurance Project Plan" (EPA, 2010). The guidance recommends assessment according to six data quality indicators (DQIs) where applicable to the data or method:

bias, representativeness, comparability, completeness, precision, and accuracy

Acceptable values and criteria for determining each DQI were defined by each TF in the QAPP section of the TQAP. Each TF was also responsible for defining reporting requirements for their DQI matrix. Both GBF and GSI identified DQIs for biology and water chemistry data sets and presented these through various combinations of matrices, tables, figures, and text. GBF also identified engineering DQIs. The DQI definitions for each facility may be summarized as follows:

GBF: Precision was the only DQI identified for biological measurements, and it was quantified by the coefficient of variation (CV), or analytical variability of the collected data. DQI objectives were based on previously observed CVs for similar analyses.

Regarding the engineering data, bias was not evaluated. Other engineering DQIs were unique to a given engineering measure, where representativeness was defined for flow as "the amount of the total flow at the monitoring point that was measured by the flow meter." This was 100% for tank and sample flow and tank

volume measures, as all flow in the piping and the total tank volume was measured using the respective sensors; a target of 90% was set for all. Comparability was measured for each of three operations per BE test through calculation of relative percent difference (RPD) of aggregate flow data at different process points, with a target of < 10% difference. Completeness was defined by comparing the number of one-minute IMAC data point intervals to the total minutes of event time for each of three operations per BE test cycle. In all cases, this resulted in 100% completeness with an objective of ≥ 90%. Precision was set at 0.1 of the parameter measured; this value was the same as the resolution of the recorded measurement. Accuracy was specified as that determined during the most recent calibration of the respective device; it is unclear how the targets were determined (possibly from device specifications).

Water chemistry DQIs were analyzed using the CV for subsamples of DOC, POC, TSS, pH, and chlorophyll *a*. DQIs were performed only for uptake samples, except for chlorophyll *a*, which was performed on "uptake and control samples which provided the highest organism densities". A qualitative comparability analysis of temperature-salinity linearity was performed for two similar but not identical locations to graphically illustrate high correlation of profiles during test periods. A graphical comparability analysis was also performed between time-integrated inline fluorometer data (chlorophyll *a* fluorescence) and acetone-extracted chlorophyll *a* from uptake water samples; this same process was used to calibrate the fluorometer. It is unclear if this was done for a single BE test or all tests. No objectives were cited for the comparability analyses.

GSI: Biological DQIs for the ≥ 50 µm and the ≥10 µm and <50 µm size classes shared similar definitions. Bias in quantification of organisms was measured through duplicate analysis by two taxonomists, and both an average percent similarity (PS, target > 90%) for abundances of living individuals in specific taxonomic groups and an average RPD (target < 20%) for total living organisms was calculated. Representativeness was defined with a qualitative objective that all samples were collected, handled, transported, and analyzed in the same manner as determined by QA/QC review. Comparability was also defined with a qualitative objective that SOPs were consistently followed as determined by QA/QC review by the QAQC Officer. Completeness was defined as the number of valid data vs. the total number of measurement samples analyzed (for that data type), with a target objective of > 90% completeness (i.e., data were deemed valid if they met all other DQI objectives for that data type). The ≥ 10 µm and < 50 µm size class also defined "precision (within sample)" that required two subsamples to be analyzed and examined for PS (target ≥ 60%) and RPD (target < 20%). Other DQIs were not analyzed for the two largest size classes.

DQIs for the < 10 µm size class used the same definitions as for the other size classes for representativeness, comparability, and completeness, but different definitions for bias were used, and some additional indicators were incorporated. Precision analysis required at least 10% of the samples analyzed in duplicate, with performance measured by average RPD (objective < 20%) of all duplicate analyses. A number of bias DQIs was identified by the TF, including:

- operator bias (RPD, objective < 20%) as defined for largest two size classes
- positive control bias (the objective was to obtain counts that were greater than limit of detection, LoD)
- negative control bias (the objective was to obtain counts less than LoD)
- method bias with procedural blank (the objective was count less than LoD)
- media blank bias (the objective was to count less than LoD)
- diluent blank bias (the objective was to count less than LoD)
- accuracy using a positive control (the objective was method-dependent)
- sensitivity, which determined the LoD for a given analytical technique.

No engineering DQIs were identified by GSI in their TQAP or presented in the verification report. Water chemistry DQIs were analyzed for precision (objective < 20% RPD), bias using blanks (the objective was method dependent), accuracy using spikes (75-125% recovery of spikes and < 20% RPD for references), and completeness (objective > 90% of samples were analyzed). Method detection limits were presented for TSS, POC, and DOC.

The DQI analysis performed by both facilities is summarized in Tables 19-21. Biology DQIs presented in Table 19 were derived from Table 8-8 in the GBF report (Golden Bear Facility, 2012) and Tables 35-37 in the GSI report (Great Ships Initiative, 2012). DQIs for ballast operations data (engineering measurements) are presented in Table 20. These were derived from Tables 8-3 and 8-4 in the GBF report. Although no DQIs for engineering parameters were identified by GSI, a review and discussion of how well operational parameters met set points was provided for each test cycle in Section 7.4 "Biological Treatment Efficacy Verification" (Great Ships Initiative, 2012), and any deviations were noted.

Water chemistry DQIs presented in Table 21 were derived from Table 8-7 and Section 8-2 of the GBF report (Golden Bear Facility, 2012) and Table 34 of the GSI report (Great Ships Initiative, 2012). The tables indicate the DQI value where applicable and indicate if they were not completely met. Those not completely met were discussed in the DQI section of the respective TF report.

Table 19. Data quality indicators for biology measurements at the test facilities.

Biology measurement	GBF DQI[A]	GSI DQI							
	Precision	Bias					Precision	Accuracy	Complete-ness
Organisms ≥ 50 µm and Organisms ≥ 10 µm and < 50 µm									
	CV	**Operator (PS/RPD)**	**Pos/Neg > LoD/ < LoD**	**Method blank < LoD**	**Media blank < LoD**	**Dilution blank < LoD**	**PS/RPD**	**Positive control**	**>90%**
Organisms ≥ 50 µm	< 31%	**> 90%**/< 20%	N/A	N/A	N/A	N/A	N/A	N/A	**> 90%**
Organisms ≥ 10 and < 50 µm (flow cytometry)	**< 42%**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Organisms ≥ 10 and < 50 µm (epifluorescence)	< 25%	> 60/ < 20%	N/A	N/A	N/A	N/A	≥ 60/< 20%	N/A	> 90%
Organisms < 10 µm									
	CV	**Operator RPD**	**Pos/Neg > LoD/ < LoD**	**Method blank < LoD**	**Media blank < LoD**	**Dilution blank < LoD**	**RPD**	**Various Pos control**	**> 90%**
Culturable, aerobic heterotrophic bacteria	< 90%	< 20%	Y/ N/A	**N**	**N**	Y	N/A	Y	**N**
Total coliforms	N/A	< 20%	Y/Y	Y	N/A	Y	**<20%**	Y	Y
Total heterotrophs	N/A	< 20%	Y/ N/A	**N**	Y	Y	**<20%**	Y	**N**
Escherichia coli	< 64%	< 20%	Y/Y	Y	N/A	Y	**<20%**	Y	Y
Enterococci	< 42%	< 20%	Y/Y	Y	N/A	Y	**<20%**	Y	Y
***Vibrio* spp. blot (DNA/RNA)**	N/A	< **20%**/<20%	**N/N**	Y/Y	Y/Y	Y/Y	**<20%**/<20%	N/A	**N/N**

[A]Values or 'N's shown in bold and in shaded cells indicate the DQI did not meet the objective in the respective test facility's test quality assurance plan. cfu = colony forming unit, CV= coefficient of variation, DQI = data quality indicator, DNA = deoxyribonucleic acid, LoD = limit of detection, N/A = not applicable, N = no, the DQI objective was not met, Operator = agreement between Analysts, Pos control = positive control, Pos/Neg = positive and negative controls (for the positive controls, the DQI was that the number of organisms detected was > the LoD, and for negative controls, the DQI was that the number of organisms counted was < the LoD), PS = percent similarity, RNA = ribonucleic acid, RPD = relative percent difference, and Y = yes, the DQI objective was met.

Table 20. Data quality indicators for engineering measurements at the Golden Bear Facility.

Engineering Measurement	GBF DQI[A]				
	Representativeness	Comparability	Completeness	Precision	Accuracy
Ballast Flow	> 90%	≤ 10%	> 90%	0.1	± 3%
Tank Level	> 90%	≤ 10%	> 90%	0.1	± 5%
Sample Flow	> 90%	≤ 10%	> 90%	0.3	± 5%

[A]All DQI objectives in the test quality assurance plan were met. DQI = data quality indicator and GBF = Golden Bear Facility.

Table 21. Data quality indicators for water chemistry measurements at the test facilities.

Water chemistry parameter	GBF DQI[A]	GSI DQI[A]				
	Precision	Precision	Bias	Accuracy		Completeness
	CV	RPD	Blanks	Spike recovery	RPD to reference	%
DOC	± 0.1 mg L^{-1}	< 20%	< 0.5 mg L^{-1}	75-125%	N/A	> 90%
POC	< 33%	< 20%	< 0.27 mg L^{-1}	75-125%	< 20%	> 90%
TSS	< 11%	< 20%	< 0.32 mg L^{-1}	N/A	< 20%	> 90%
pH	± 0.1	< 20%	N/A	N/A	N/A	N/A
Chl *a*	< 13%	N/A	N/A	N/A	N/A	N/A
% T	N/A	< 20%	> 98% T	N/A	N/A	> 90%

[A]All DQI objectives in the test quality assurance plan were met. Chl *a* = chlorophyll *a*, CV = coefficient of variation, DOC = dissolved organic carbon, DQI = data quality indicator, GBF = Golden Bear Facility, GSI = Great Ships Initiative, N/A = not applicable, RPD = relative percent difference, TSS = total suspended solids, and % T = percent transmittance.

6.4.2 Comparison of Biological DQIs

For biological parameters, GBF used a DQI for precision (CV), and, in most cases, met the target objective (Table 19). At GSI, operator bias and completeness were assessed for the two largest size classes and precision for the ≥ 10 µm and < 50 µm size class was assessed (Table 19). Precision analysis using the CV was not provided for the ≥ 50 µm size class, although this measure was cited in the GSI QAPP. Regarding the data quality objectives for bias and completeness in the ≥ 50 µm size class, they were met for the macrozooplankton fraction, but not for the microzooplankton fraction of the size class (that level of detail is not shown in Table 19). All GSI objectives were met for the ≥ 10 µm and < 50 µm size class.

For organisms < 10 µm, GBF used precision (CV), and GSI provided a suite of DQIs, consisting of several methods to assess bias using positive and negative controls and media and process blanks. GSI also assessed operator bias by comparing results from multiple Analysts. Completeness, precision, and accuracy objectives were also assessed. Precision objectives for this size class were not met in most cases. In addition, one negative control for *Vibrio spp.* showed positive growth, and several blanks for cultured heterotrophic bacteria and one for total heterotrophic bacteria showed growth. These missed objectives also resulted in the associated completeness objectives not being met. In their verification reports, both TFs discussed the DQIs identified in their respective TQAPs, with the single exception for GSI noted above (precision for the ≥ 50 µm size class).

6.4.3 Comparison of Engineering DQIs

In their discussion of engineering parameters, neither facility presented any DQI analysis for bias, and this DQI might best be separately addressed as part of an overall facility validation test not involving a treatment system. Regarding other engineering DQIs, the ETV Protocol states, "representativeness entails collecting a sufficient quantity of data during operation to be able to detect a change in operations" (EPA, 2010). Both facilities presented data that demonstrated this capability through graphs and histograms in a discussion of operational results, although neither presented it under the DQI discussion. A thorough analysis of flow comparability across all tests was provided by GBF, in which the summation of flows matched the primary flow measure, and all objectives for ballast and sample flow were met (Table 20).

Regarding engineering DQI objectives for completeness, by definition GBF met all of them; the DQIs were defined as the points collected by their automation system. GSI did not provide any objectives or analysis for comparability and completeness. Precision was cited by GBF as the value recorded (i.e., the objective was precision of 0.1, and data were recorded to 0.1 units) and thus met the stated objective. Neither TF provided any analysis per the ETV definition, "Precision refers to the mutual agreement among individual measurements and provides an estimate of random error" (EPA, 2010).

Accuracy for engineering parameters was addressed by both facilities by calibrating sensors, which was discussed by GBF under the DQI section and GSI under the QA/QC discussion. GSI provided tables with calibration dates for key sensors. Since all sensors were calibrated, both TFs met the accuracy objectives. GBF provided analysis for all the DQIs cited in their TQAP. It should be noted that under the TQAP approved by the VO, GSI did not specify any DQI analysis for operational parameters, so the lack of operational DQI discussion cannot be considered an omission.

6.4.4 Comparison of Water Chemistry DQIs

The DQI analysis of water chemistry parameters was reported using precision measures by both TFs, wherein GBF assessed the CV and GSI assessed relative percent difference for duplicate analyses (Table 21). GSI also provided DQIs for bias, accuracy, and completeness. The latter were assessed with blanks, spikes, and a review of samples collected vs. samples analyzed. Both GBF and GSI met all their objectives for quality of water data.

6.4.5 NSF International Audits

The AO for this project was NSF, whose experts reviewed standard operating procedures and TQAPs for both TFs, were present for at least one BE test at each facility and performed a full audit of GSI prior to the BE tests. GBF could not accommodate a full audit of the facility before the BE tests; therefore, NSF reviewed the available documentation and SOPs and provided comments to GBF. Because NSF is producing a written report as part of this project regarding the finding of their audit and reviews, this report does not discuss the detailed audit findings but instead provides a general overview.

NSF produced a list of comments regarding the review of the GBF facility operations and test data after all tests were completed at GBF. The comments noted the lack of training records, instrument calibration data, equipment records, traceability of data, and updates to the data sheets or SOPs. The NSF comments regarding water quality highlighted the need to check the calibration of instruments against calibration standards and encouraged GBF to obtain QA/QC information from laboratories that conducted analyses by subcontract to GBF. NSF also identified an issue with GBF's TSS procedure, noting that it did not follow the recommended ETV methods. In addition, the adenosine triphosphate (ATP) analysis method cited in GBF's raw data notebook differed from the method in the written SOP. Finally, NSF identified several issues related to GBF's biological analysis and documentation procedures: a lack of records, missing or complete SOPs for maintaining records, lack of traceability for media preparation, and the absence of a spike/recovery study for the Colilert® and Enterolert® analyses. NSF noted a need for better quality control for HPC data, as the incubation temperatures were outside of specifications, and there were discrepancies between the raw data counts and data entry.

An important comment was related to the Colilert® review. During the analysis of the uptake samples for BE test 1, the temperature of the incubator rose to 57°C for ~2 h. NSF indicated this temperature exceeded the limit of the assay; therefore, the results were invalid. In the GBF verification report, the invalidity of the data was not mentioned. Unfortunately, GBF did not have a chance to make corrective action to the procedures during the BE tests because the review occurred several months after the tests had been completed. Many of these issues likely would have been discovered and addressed if an audit had been conducted prior to testing.

NSF conducted a complete audit of the water quality and biological procedures performed at GSI. The audit was conducted prior to testing, and the comments provided by NSF were used by GSI during the BE tests. The auditors reviewed the GSI SOPs to ensure the procedures were complete and followed with proper quality control throughout the test period. In general, the audit of water quality procedures found relatively small issues such as missing or incomplete chain-of-custody forms, documentation that was missing, or SOPs that were not followed exactly as written. The biological review audit indicated issues with data entry and data validation for the database for the ≥ 50 µm data. Other observations indicated changes were needed for the data sheets for HPC, and additional checks were needed during processing. For all items found in the audit, GSI responded to NSF with corrective action and made changes as needed.

6.4.6 Staffing

Each TF identified a set of personnel and contractors participating in the Intercomparison Project. The personnel included those involved in developing the project documentation as well as developing the facility and experimental designs, but not all personnel were involved in the physical testing of the BWMS. The staff working at each TF was consistent with the facility's respective proposal and subsequent TQAP, both in terms of roles and qualifications. In the GBF verification report (Golden Bear Facility, 2012), 19 participants were identified: 5 from MLML, 6 from CMA, 7 CMA student assistants, and 1 from The Glosten Associates. Additional assistance was provided by the subcontracting organizations McCampbell Analytical (which conducted water chemistry analysis) and Nautical Engineering (which provided marine engineering services during installation and commissioning of the BWMS). In the GSI TQAP, 18 people were identified: 3 from the Northeast Midwest Institute (NEMWI), 2 from AMI Engineering, 1 independent consultant, 9 from Lake Superior Research Institute (LSRI), and 3 from the Natural Resources Research Institute (NRRI). AMI Engineering also provided engineering services, additional staffing during installation of the BWMS, and an engineering student intern during commissioning and testing.

During commissioning and BE testing, the number of staff dedicated to testing was one of the more notable differences between the two TFs. At GBF, a few key staff members were responsible for multiple functions and thus wore "multiple hats", particularly during testing. Additionally, members of the Science Team (the Science Team Lead and Analysts) were directly responsible for manually operating the sample collection apparatus as well as carrying out their science duties. In contrast, at GSI, most teams had multiple members and the overall on-site project manager was not directly responsible for any hands-on tasks during test operations; thus, he monitored testing and assisted as needed. For sampling operations that were not automated, the sampling tasks were performed by dedicated staff, mostly from the laboratory at LSRI, who thus augmented the on-site science team.

Table 22 below shows the staffing observed during BE tests at each TF (note that these staffing arrangements were observed during BE tests, and they differ slightly from the numbers in the TFs' TQAPs). It also lists additional staff who were available but not necessarily on site during a test. Each facility identified staff using different job titles, but to allow a side-by-side comparison between TFs, the job titles used at the two facilities have been mapped to the 'functions' column in the table. The functions in the table are defined as follows:

- The Facility Director was responsible for the proposal to participate in the Intercomparison Project, the facility, and subsequent execution of the project at the respective TF
- The Project Manager was responsible for managing the overall TF activities on a daily basis and he or she coordinated the various teams associated with the Intercomparison Project, both on- and off-site
- The Test Lead Operator was responsible for operating the TF; he and his team (test operation staff) performed the engineering aspects of testing and TF operations such as equipment handling, ballast operations, BWMS monitoring, in-line instrumentation, automated sample collection, and facility pre- and post-test grooming
- The Sampling Staff were responsible for the collection and chain-of-custody for samples in each biological size class and the whole water samples for chemical analyses. Any sample transport was also managed by this team
- The Science Team was headed by one or more Lead Scientists, and team members (Scientists and Analysts) performed all sample analyses. Analyses for the two largest size classes of organisms

were performed on site; analyses for the smallest size class of organisms and for water quality were primarily performed off-site at team or contractor laboratories by both TFs

- QA/QC staff (Auditor and Database Manager) collected, audited and logged data sheets, documented redlines or changes, maintained calibration records, and performed checks of data entry and analytical results

Table 22. Test facility staffing observed during testing for the Intercomparison Project.

Function	GBF TQAP Job Title		GSI TQAP Job Title	
	Test-Day Personnel	**Other Personnel**	**Test-Day Personnel**	**Other Personnel**
Facility Director	1 Facility Director	-	-	1 Principal Investigator
Project Manager	Person in Charge (the Facility Director)	-	1 Facility Site Manager	-
Test Lead Operator	Lead Operator (also the Facility Director)	-	1 Operations Manager	-
Test Operations Staff	Typically 2 Ship's crew and up to 4 Student Assistants	3 Ship's Crew	1 Staff Engineer 1 Intern	-
Sampling Operations Staff	4 Sampling Staff (performed by Science Team)	-	4 Sampling Staff (included some Science Team members listed under 'Other Personnel')	-
Science Team Lead	1 Science Team Lead	-	2 Team Leads	-
Scientists and Analysts	3 Analysts	-	3 Zooplankton Analysts	1 Zooplankton Analyst[A]
	-		2 Phytoplankton Analysts	1 Phytoplankton Analyst
	-	1 Backup Analyst	-	2 Microbial Analysts
	-	Chemistry subcontractor (McCampbell Labs)	-	2 Chemists
QA / QC Auditor	1 QA Officer	-	2 QA/QC Officers	-
Database Manager	-	DB Manager (also the QA Officer)	-	1 DB Specialist

[A]One microbial Analyst could also provide support as an additional zooplankton Analyst. DB = database GBF = Golden Bear Facility, GSI = Great Ships Initiative, QA = Quality Assurance, QC = Quality Control, and '-' = no staff member.

6.4.6.1 Staffing—Golden Bear Facility

In the following discussion, personnel performing job functions are identified by the job title used by the TF in their TQAP. In Table 22, any job function fulfilled by the holder of another job function is identified in parenthesis. This format shows instances of a person filling multiple roles. At GBF, this arrangement was not uncommon: the Person in Charge was also the Lead Operator; both functions were performed by the GBF Facility Director. The Lead Operator coordinated actions using a computer-based communications method to direct and monitor commands to the pump and valve operators in shaft alley, the QA Officer in

the IT room, the sampling team on the main deck, and the BWMS on the 01 deck. He also tracked operations using GBF's IMAC interface. As the Person in Charge, the same person physically monitored activities of each of the teams, travelling to the various locations about the ship to ensure all processes were operating correctly and addressing the operational issues of the BWMS, that is, the water hammer and need to conduct manual backflushing.

The GBF Science Team, which was augmented with Student Assistants, performed all duties associated with sample collection under direction of the Science Team Lead, and then processed and analyzed samples in the shipboard laboratory. This team prepared and analyzed all samples for the three size classes and also performed ancillary analyses (e.g., measuring ATP and using flow cytometry to quantify organisms in the ≥ 10 µm and < 50 µm size class). They also transported water samples to the off-site (subcontractor) laboratory.

The QA Officer monitored the in-line water chemistry instrumentation, automated data collection, and instrument calibrations. He also processed hand-logged data and occasionally assisted with sampling operations. Due to these multiple duties, the independent QA oversight during testing was minimal. Subsequent to testing, the QA Officer performed data audits and verified data entry. He was also responsible for maintaining the database and backing up data. Because GBF is at a maritime academy, on occasion, a CMA staff member was unavailable during testing due to teaching responsibilities. In this case, an alternate CMA crewmember would fill in or one of the other staff would assume those duties.

6.4.6.2 Staffing—Great Ships Initiative

At GSI, the Facility Manager acted as a facilitator and communications director, coordinating activities with the various teams both on- and off-site, and he was also in regular communication with the Principal Investigator, who was off-site for two of the BE tests. The Facility Manager had no specific assignments in daily test operations, so he was available to help as needed. The Operations Manager initiated and monitored test activities, and all water sampling, except for whole water samples, was automated through the GSI HMI console. In practice, at least one of the QA/QC Officers coordinated and logged activities of the sampling team and audited activities while they occurred. Calibration records, SOP monitoring, and data audits were shared between the QA/QC Officers. On-site QA/QC monitoring was visible during test operations. Sampling was performed by a team comprised of on and (normally) off-site personnel. The latter came on-site to support specific sampling activities and then transported water samples back to the off-site laboratory for subsequent analyses (primarily organisms < 10 µm and water chemistry). The on-site personnel collected and provided samples to the on-site laboratories and then commenced post-test equipment grooming. At GSI, one Team Lead was responsible for the activities of the zooplankton Analysts, and another Team Lead was responsible for the phytoplankton Analysts. Both analyses were performed in separate facilities on site; these groups also generated ambient counts immediately prior to running a test. Additional personnel were available from the university laboratories to staff test operations as needed if someone were sick, support plankton tow harvesting, etc. The database was managed off-site by a dedicated IT database specialist. Science team and engineering data were logged to the database, and audit reviews were performed by the QA/QC Officers.

6.4.7 Effectiveness of QA/QC Practices

Various QA/QC procedures and metrics were employed by both TFs to address requirements of the ETV Protocol. Effective QA/QC practices consider data variability, overall adherence to measurement protocols, and the statistical power or accuracy and precision of the measurement method. Adherence to protocols and

accuracy and precision for sensors was assessed using audits, records, and logs. In addition, the use of DQIs was specified by the ETV Protocol to quantify confidence in data obtained during testing, with specific selection left to the TF. DQI effectiveness is determined by selection of appropriate indicators and objectives that can highlight abnormal variance or ranges in values for the data and measurement method.

Manual counting of live organisms was assessed for variability consistent with the statistical distribution of organisms, meaning that, ideally, a TF also provided a level of confidence associated with the stated measurement. Tests that examined operator bias and comparability provided further confidence in the resulting data. The use of blanks, spikes, or other positive and negative controls consistent with population assessment methods provided confidence that execution of the method was effective for each test. Here, completeness was also an effective assessment of data quality.

Engineering indicators and objectives provided confidence in assessing performance and variability with respect to set point or operating ranges. Many engineering parameters are continuously monitored through logging of in-line sensor data, so the use of completeness was generally a less useful indicator than data showing adherence to a set point or operating range. In-line sensors, when properly maintained and calibrated, provide quantifiable accuracy and precision values for a measurement. Thus, a record of calibration for the sensors was a primary QA/QC attribute. Post-test comparisons of a given parameter using multiple measurements to corroborate data (e.g., comparability between individual sub-flows and a measured total flow) provided additional confidence in the data that previously calibrated sensors provide, and can give additional insight into the effective precision and accuracy of measurements within the TF system. Representativeness offered a qualitative assessment of consistency in data sampling, handling, or measurement as monitored by QA/QC personnel also provided confidence in data quality, particularly when the method is dependent on multi-step processing of samples.

Tracking of deviations and changes to the SOPs was most effectively recorded and managed when an independent QA/QC officer was able to log and document each incident, and it resulted in more accurate reporting of deviations in the TF verification report. Redline markups of SOP log sheets by multiple staff members were less effective, especially if there were no data recorded, or a scan log was incomplete. Use of pre-printed data sheets based on previously vetted SOPs enhanced the ability of the VO and auditors to review test data, as did well-organized binders for those items not stored electronically, such as training, maintenance, or calibration logs.

Staffing appeared to have a direct effect on the ability of a TF to implement QA/QC processes and maintain well-organized records. A single individual who performed multiple roles in the testing process allowed important or time-critical decisions to be made by a person with direct knowledge informed by those roles. However, given the multiple responsibilities, the same individual has less time for documentation of event details, logging of quality metrics, or responding to simultaneous events. Alternatively, when professional staff (rather than students) conducted the various activities involved in a test, higher quality (fewer errors, as determined by NSF auditing) and more detailed data were recorded, and senior staff were available to oversee actions and respond where needed. The ability to devote QA/QC staff to monitor, assist and record observations, but not be committed to test roles improved confidence in the quality of the test data, and the use of QA/QC personnel to independently assess data transfer and analytical calculations improved the confidence in the reliability of the reported results.

Finally, the more effective QA/QC systems included an intrinsic process for continuous self-improvement, as evidenced through internal and external audits, corrective and preventive actions, and management

review. These mechanisms for documenting effectiveness of QA/QC practices allow for objective feedback and incorporation of constructive criticism into standardized testing, staff training, and management processes.

6.5 Data Management

The ETV Protocol describes several procedures relating to data management in Chapter 8 "Data Management, Analysis and Preparation" (EPA, 2010). The first part of the section pertains to data entry during testing activities. Data entry requirements and the TFs' adherence to them are summarized in Table 23. All requirements were met, except for the documentation of some deviations by GBF.

Table 23. Data management, analysis, and presentation requirements and adherence to them by the test facilities.

ETV Protocol data management, Analysis, or Presentation Requirement	GBF (Y/N)[A]	GSI (Y/N)
Data are entered directly, promptly, and legibly	Y	Y
Hand-entered data are recorded legibly in ink; all original data records include, as appropriate, a description of the data collected, the unit, the unique sample identification, the name of the person collecting the data, and the date and time of data collection	Y	Y
Any changes to the original entry do not obscure the original entry, document the reason for the change, and are initialed and dated by the person making the change	Y	Y
All deviations from the QAPP must be documented in writing, and approved by the TF; documentation and communication include an assessment of the impact the deviation has on data quality	**N**	Y
Data in electronic format shall be included in a commercially available program for word processing, spreadsheet calculations, database processing, or commercial software developed especially for the data collection and processing on a specific hardware instrument or piece of equipment; backup of computer databases should be performed on a daily basis, if possible	Y	Y

[A]Values shown in bold font and in shaded cells indicate the requirement was not met. ETV = Environmental Technology Verification program, GBF = Golden Bear Facility, GSI = Great Ships Initiative, QAPP = Quality Assurance Project Plan, TF = Test Facility, and Y/N = yes/no.

6.5.1 Golden Bear Facility Data Management

GBF used notebooks and pre-printed data sheets (collectively referred to "hand logs"), to keep track of hand-written data during BE tests. The hand logs were filled out during the BE test and collected daily by the Quality Officer, who scanned the logs into GBF's online information system. GBF's Quality Officer initialed all hand logs, indicating he had reviewed the data to confirm all writing was legible and all complete. Automatically generated data—including valve positions, system water flow rates, sample flow rates, system pressures, temperature, water level within the control and treatment tanks, sample volume, power quality, main ballast pump status, and BWMS status—were tracked and recorded by IMAC.

For any deviations from the written procedures, GBF used a "redline" procedure. The person in charge of the operation made the redline change to the procedure using indelible ink by inserting new text and striking a line through the text to be deleted. The person's initials were written next to the change, and the change was explained in the notes section of the SOP. There were several instances during the GBF testing period where changes were needed and the redline procedure was used. The procedures were then reviewed by the

Quality Officer. GBF discussed in the verification report many of the deviations (with the exception of the last two deviations identified in Table 4) that occurred during the BE tests.

The IMAC data were exported to Microsoft® Excel from the operator interface terminal. Hand-written data, such as information contained in biological notebooks, were transferred manually from the data sheet to Excel. An SOP for transcribing data from notebooks to an electronic spreadsheet was not included in the documents submitted by GBF. The name of the person entering data, the date of entry, and indication that the data were checked was not included in the Excel spreadsheets. After the data were entered, they were printed in printable document format (.pdf) then sent as part of the verification report. All electronic data were stored on a hard drive under control of the Lead Scientist. The specific procedures of backing up the hard drive and the frequency of backup were not written in the GBF SOPs, although a procedure was discussed during the initial review of the facility by the VO. A general description for records maintenance was included in the GBF Quality Management Plan (QMP) that indicated all documents and records were to be scanned to a central system for a period of five years. (Golden Bear Facility, 2010). GBF shared electronic data using a file transfer protocol server that was owned and managed by The Glosten Associates. This file server was password protected, with access only allowed for participants in this project. The server was not mentioned in the GBF TQAP or QMP, so it was unclear if this server was used to store all information obtained during the testing at GBF.

The engineering data submitted by GBF to accompany the verification report included a top-level table summarizing the contents that identified the SOP forms (Appendix A), or Data Plots/Tables (Appendix B) for each Test Cycle, but when reviewing the documents submitted, the naming system for the files was not always the same as listed in the table. In addition, some files appeared as though the file name was missing a revision number. The processes used for file naming and locating files were often difficult for the reviewers to follow; typically page numbers were not provided for the data within a Test Cycle, while other data files had no index or table of contents, only scanned pages in what appear to be chronological order.

6.5.2 Great Ships Initiative Data Management

The data for water quality, engineering, and biological sample analysis were recorded on pre-printed forms or laboratory notebooks. The laboratory notebooks were numbered and identified with a code specific to this project. After a test was complete, the data collections forms and notebooks were scanned and converted to printable document files (.pdf) immediately after completion of the test cycle. The original data collection forms were stored in three-ring binders identified with a code specific to the Intercomparison Project. The electronic files were stored on GSI's internal Microsoft® SharePoint® site. The SharePoint® site was password protected with access only for personnel involved in this project. In addition, the GSI archived all original raw data in a climate-controlled, secure room at LSRI where it will be kept for a period of at least seven years.

GSI described the process for entering data from the data collection forms and notebooks in their verification report (Great Ships Initiative, 2012). Water quality data were entered directly onto Microsoft® Excel spreadsheets, and 100% of the data entry was verified by another individual who did not enter the data. All cells containing formulas were locked and GSI had a written SOP for entering the raw data into individual worksheets. For data pertaining to organisms ≥ 50 µm, GSI entered data from the form into a Microsoft® Access Database that was managed by the GSI Database Manager ("GSI Zooplankton Database"). The density of living organisms for this size class was calculated automatically, and 100% of the data entry was verified against the original data by the GSI Senior QA/QC Officer. Hand calculations of organism densities were verified against the database results for approximately 10-25% of the data. For

data in the ≥ 10 µm and < 50 µm size class, GSI used another database ("GSI Phytoplankton Database"). Again, the entered data were checked against the raw data in its entirety. Manual calculations were done for 10% of the data to verify the densities of organisms calculated by the database. A similar database was developed for the organisms < 10 µm. All data entered into the database were verified against the raw data. A similar process occurred for all engineering data. During this project, GSI did not have a procedure for managing raw microbial data, data entry, or data analysis. After each test cycle was complete, GSI determined if each BE test was valid based on the original test validity criteria detailed in the TQAP.

Deviations were managed by communicating the deviation to the GSI QA/QC Officers. A "GSI SOP Deviation Form" (in the GSI QAPP) was completed, including the date, time, description, and any impact the deviation might have on the test (Great Ships Initiative, 2010). The deviation form was then signed by the PI, and deviations were reported in the final report.

GSI allowed access to the SharePoint® site to obtain the appendices of their report. A list of files contained in the appendices was provided, and instructions to navigate through the SharePoint® site were provided. The instructions allowed reviewers to match all files with the files listed in the verification report.

6.5.3 Effect of Data Management Approaches

Both GBF and GSI generated and logged data from a variety of sources, some electronic and some by hand. Each facility opted to convert the majority of hand-logged data to electronic storage. This was generally performed by directly scanning the hand-logged files into a computer file or by transcribing data from hand logs into a spreadsheet. To be effective, electronic data management requires that data are well organized at the time of origin, stored with naming and file structures that are intuitive, and are accessible while also having protections in place to prevent unintended (or unauthorized) modification. Such a system also enhances QA/QC processes—such as transcribing data and checking analytical calculations—to be conducted in an efficient manner. No provision for version control was incorporated in the file naming system by either TF; a means for identifying revisions within file names would benefit both test facilities and reviewers.

Based on observations and review of data and analytical processes throughout the testing period, the most effective use of hand-logged data was (1) using pre-printed log sheets or data entry forms specific to the task or event at hand, and (2) using forms designed to capture information for no more than a single day. Both data and QA/QC records seemed more complete when technical staff involved in test activities were responsible for completing only data entry, and tracking of deviations or redlines to procedures was recorded by a member of the QA/QC staff. This process allowed technical staff to alert QA/QC staff and then concentrate on test activities, while the QA/QC staff proceeded to document the information needed to track and respond to the issue, consulting as needed with appropriate technical staff. This approach also provided the VO reviewers the most transparent means of following test data from generation to results and resulted in more consistent and complete tracking of test deviations and redlines in the verification reports.

Electronic data storage requires establishment of data structures or databases with associated computer access, security and backup procedures for reliable and effective use. Clearly defined (and enforced) methods to allow authorized data entry and retrieval, which also prevent unauthorized or accidental changes to previously validated computer analysis procedures, provided additional confidence in results. In general, when validated spreadsheets were used, organization of data was cleaner and easier for reviewers to follow, especially when auditing against hand-logged results. Having a well-defined and established means for remote access to TF data also aided review of results between the VO and TFs. Minimal details were

provided by either TF on how electronic data from the TF data acquisition and control system were handled in terms of transfer to a facility database or its security and backup procedures; ideally, this will be addressed in future facility QAPPs.

Finally, the establishment and upkeep of both data management and QA/QC systems is an ongoing process. Such processes continue to improve when they include a means to observe and track deviations and issues, and then incorporate those observations to improve future practices. This documentation and periodic review and update of systems are inherent (and time-consuming) components of effective quality and data management systems. Documentation is also a key requirement for standardized testing–auditing practices dictate that if an action was not documented, it was not performed. Both TFs have established means to incorporate detailed test documentation and the process of continuous improvement; future testing should benefit from these actions.

6.6 Verification Reporting

A primary output from the EPA's ETV Program is the verification report, which documents the process and results of testing. A companion output is the shorter verification statement, which summarizes, at a top level, the test conditions and results. Participants in the Intercomparison Project were required to provide a verification report consistent with ETV requirements but were not required to produce a verification statement, as this ETV document is generally used by vendors for commercial purposes, which was not the intent of this research project. The following definition was provided to the TFs, consistent with that in the ETV Protocol (EPA, 2010):

> "Verification Report: A detailed report on the testing results of a particular technology according to an approved Test /Quality Assurance Plan and conducted under the ETV Program. The report is typically prepared by the TO and contains a description of the test facility, photographs of technology being tested, methods and procedures, presentation of analyzed data including all QA/QC data obtained during the test. Appendices include raw data sets and lab audit information, TQAP, O&M Manual and other relevant information. Both the verification report and verification statement are publically available on the ETV Program's web site and NSF's web site."

As the Intercomparison Project was not conducted through the EPA's WQPC, the verification reports were reviewed by the VO and AO. As mentioned earlier, the two TF verification reports resulting from the Intercomparison Project are included in the References section of this report, and the reports' contents are listed in separate appendices (Appendices A and B) to this report. Publication of the Intercomparison report and the TFs' verification reports are under the jurisdiction of the USCG as the funding agency, not the EPA. Thus, these reports might not be publicly available on the EPA ETV web site, and the TFs were notified that any public dissemination of results was subject to USCG authorization.

The guidance for reporting in the ETV Protocol is provided in Chapter 6 "Reporting Verification Testing Results" (EPA 2010), and states "Deviations from this protocol or any TQAP prepared for BWTS [Ballast Water Treatment System] testing shall be described in the verification report, which shall include supporting documentation that provided the basis for acceptance of the deviations. All testing results will be presented in the report, including all data regarding challenge conditions, results of verification testing for all verification factors, and any vendor-supplied data or information. A summary verification statement will also be prepared." Both TFs reported deviations that occurred during testing and included a discussion of any consequences resulting from the deviations. As part of their report, GSI included an appendix

documenting their various audits and addressing findings according to the ISO 9001 categories of "non-conformance", "deviations", "observations", and "opportunity for improvement".

The ETV Protocol also provides a list of sections and appendices to be included in the verification report. Additional instruction from was given to the facilities to include the overall BE test validation matrix in the executive summary. The required content for the Intercomparison report is summarized in Table 24 below, which lists the mandatory sections and each TF's corresponding chapter names per their respective table of contents. The table identifies how the submitted verification reports follow the requirements of the ETV; in general, the body of the reports closely followed the outline of the ETV requirements. The differences among titles and report organization illustrate that additional guidance, such as a draft verification report, template, or series of checklists would help standardize reporting among TFs, ensure reports are as comprehensive as desired, and allow easy comparison among reports.

Table 24. Requirements for verification reports and the contents of the test facilities' reports.

Verification report requirements for Intercomparison Project	GBF verification report TOC	GSI verification report TOC
Executive Summary (to include BE test validation matrix)	Executive Summary with Validation Matrix	Executive Summary with Validation Matrix Acknowledgements
Introduction and background	Project Introduction and Background	Introduction The Testing Organization and Testing Facility
Description of the treatment system	Ballast Water Management System Description	The Ballast Water Treatment System
Experimental design (including a description of all deviations from the protocol and the basis for accepting the deviations)	Experimental Design	Test Objectives and Experimental Design
Description of challenge conditions	Challenge Water Conditions	Challenge Conditions and Preparation
Methods and procedures	Methods and Procedures	Methods and Procedures
Results and discussion	Scientific Results – Biological Efficacy	Results
Verification testing operation and monitoring QA/QC	Quality Assurance and Quality Control Data Quality Indicators GBF ETV Conclusions Acknowledgements	Verification Testing Operation and Monitoring Quality Assurance and Quality Control Discussion of Results Conclusion References
Appendices	Appendices A-H	Appendices 1-4
TQAP	Appx A: Project TQAP with Redlines (Redacted)	Appx 1: GSI TQAP with 8 appendices
Vendor-supplied Operation and Maintenance Manual	Appx A includes vendor operations manual as Volume IV in non-redacted TQAP	Appx 1 includes vendor operations manual as TQAP Appendix 1
Data generated during testing	Appx B: Data Logs – Engineering (181 pages of scanned forms and logs) Appx C: Data Logs – Automation System Outputs (551 pages of data plots, truth tables and raw data) Appx D: Data Logs – Water Quality Analysis Reports (30 pages scanned logs, primarily from McCampbell Analytical) Appx E: Data Logs – Biology Notebooks (125 pages scanned notes)	Appx 2: Commissioning Acceptance Form_ Completed_ Redacted Appx 2: Data Generated During Installation, Commissioning, and BE Testing (25 raw data spreadsheet files for Operational, BE and Water Quality) Appx 3: Detailed Measurement / Sizing Data from BE Verification (2 raw data spreadsheet files)

Table 24. Requirements for verification reports and the contents of the test facilities' reports (Continued).

Verification report requirements for Intercomparison Project	GBF verification report TOC	GSI verification report TOC
QA/QC records	Appx F: QA/QC Data and Figures – Biology Efficacy Raw Data Listings (61 pages tables and charts) Appx G: QA/QC Data and Figures – Engineering Parameters (65 pages data plots) Appx H: QA/QC Data and Figures – Water Quality Parameters (18 pages data plots)	Appx 2: Data Generated During Installation, Commissioning, and BE Testing (4 raw data spreadsheet files for DQOs) Appx 4: Technical Systems Audit Findings (report)
Maintenance logs	None supplied	Logs reviewed on site during testing by VO
Any other records maintained during testing, such as chain-of-custody forms	NS	NS
Any other information provided by the Vendor, which may be of use to the stakeholder community	NS	NS

Appx = appendix, BE = biological efficacy, ETV = Environmental Technology Verification program, GBF = Golden Bear Facility, GSI = Great Ships Initiative, NS = not submitted, QA/QC = quality assurance/Quality Control, TOC = table of contents, and TQAP = test quality assurance plan.

The appendices to the verification reports contain data and information generated during the project, and here, some differences may be observed in the type of information provided by each TF. In reviewing the raw data submitted, the data in spreadsheet form (rather than .pdf) were easier to audit where calculations are involved, as formulas can be examined. In general, spreadsheets that included tabbed page names and used color within each spreadsheet page made it easy to identify specific test data and follow corresponding QA/QC analysis. It is recommended that future verification reports include data in this format. It is also important that the verification report reference all data in the appendices (i.e., by providing direction to the specific section within a given appendix) to provide supporting data for results and conclusions; in some instances, data in appendices were not referenced in the report, which allows them to be overlooked or misinterpreted. Both facilities did provide access to scanned raw data, which was used to check random samples of the data transcription to spreadsheet. Neither TF provided calibration or maintenance logs as part of the report, but during BE testing at GSI, calibration, maintenance and testing logs were requested and reviewed by the VO. GBF provided chain-of-custody records from McCampbell Laboratory as part of the scanned Water Quality documentation. Neither facility provided internal chain-of-custody documentation. Given the volume of information presented in the verification reports, it is not surprising that information not specifically required was not presented. Additional guidance within the ETV Protocol to define organization and presentation of audit reports as well as raw and QA/QC data within appendices would be beneficial, where audits include the review of calibration, training, and maintenance logs.

Preparation of the verification reports was a time-consuming process and took longer than anticipated by both the TFs and the VO. Neither TF had produced such a document nor had previously conducted testing according to the rigors of the ETV Protocol. In particular, the lack of guidance within the ETV Protocol on DQIs and objectives led to different interpretation between the participating facilities. Published verification reports for other ETV protocols provide additional examples and are available on the EPA ETV WQPC website http://www.epa.gov/nrmrl/std/etv/vrvs.html. The VO, in conjunction with the AO, provided reviews with editorial suggestions to each draft verification report. During this review and revision process, it became apparent that document revision control and naming conventions should be improved at both facilities.

7 SUGGESTED CHANGES TO THE ENVIRONMENTAL TECHNOLOGY VERIFICATION PROTOCOL AND COMPARISON TO THE INTERNATIONAL MARITIME ORGANIZATION G8 GUIDELINES

A secondary aim of the Intercomparison Project was to make recommendations, if needed, to improve the ETV Protocol. Indeed, as the TFs prepared for and conducted BE tests, a number of questions and issues were raised. Many of these issues deserve consideration by the ETV Technical Panel for future revisions to the ETV Protocol. Both of the TFs and the VO made lists of the concerns and provided suggestions to ameliorate them.

7.1 Suggestions by Test Facilities to Improve the Protocol

The TFs' suggestions are discussed in depth in their individual verification reports and can be read in their entirety in Section 9 "Golden Bear Facility ETV Conclusions" (Golden Bear Facility, 2012) and Section 9.3 "The ETV Protocol" (Great Ships Initiative, 2012). In general, the suggestions can be grouped under the broad headings of Experimental Design and Conditions, Interaction with the BWMS Vendor and Class Society, Sample Collection and Analysis, and Guidance for the verification report. Table 25 shows the TFs' suggestions grouped by these headings, a summary of their recommended actions, and the VO's feedback to the recommendations. The TFs, having run BE tests according to the ETV Protocol, were well positioned to make recommendations that will spur discussion and needed research. In nearly all cases, ESAC and the TA agreed with the suggestions or agreed they warranted further examination.

Table 25. Suggested changes to the Environmental Technology Verification Protocol.

Type of suggestion	TF(s)	Recommendation by TF	Comment by the VO
Experimental design and conditions	GSI	Increase the number of BE tests (i.e., from 3 per salinity level to 5) to increase confidence in the data.	While it is generally desirable to have greater replication, the cost-benefit relationship may not support it in this case. Data should be collected on the utility of additional tests, and if higher replication proves beneficial, the ETV Protocol should be changed accordingly.
	GSI	Fill the control and treatment tanks simultaneously (vs. sequentially) to ensure the communities of organisms are similar in species composition and density.	If possible, tanks should be filled simultaneously, and new TFs should be constructed in this manner.
	GSI	Reconsider the requirements for challenge conditions, e.g., to meet the POC requirements, GSI must augment ambient water, which also increases DOC, which is already high at GSI.	The challenge water conditions should be revisited by the ETV Technical Panel to determine their relevance to real-world conditions and their practicability; in this project, none of the tests met all ETV challenge water requirements as listed in the ETV Protocol due to various facility and geographic constraints. Likewise, the validation that should be done when augmentation of POC, DOC, and MM is used should be discussed, including the potential need to repeat the validation for biological communities that change seasonally at TFs.
	GSI	Determine if the O&M guidance in the ETV Protocol is appropriate, that is, conduct an Intercomparison test for O&M testing.	A study of the O&M testing in the ETV Protocol might be useful to determine the need for additional Intercomparison testing, or O&M testing might be accomplished during the shipboard testing required for U.S. Type Approval.
Interaction with BWMS vendor and class society	GBF	Include the vendor in the development of the test plan (as per the ETV Protocol) as well as 'equipment set-up and commissioning support, test cycle support, test cycle preparation and initiation witness, review of preliminary data and analysis, and participation in any investigation if indicated'; suggest guidance to be included in the ETV Protocol if a vendor's claim is not met.	Because ETV testing is different from this Intercomparison research project, the vendor's role should be expanded relative to the work done here, and sharing the data with him or her is appropriate. Nonetheless, the TF must remain independent from the vendor, and if changes were made to the BWMS during the course of an investigation, the tests would need to start over from the beginning.
	GBF	Enlist a classification society to verify the BWMS is installed properly and provide operating parameters and limits for the TF to monitor and verify during testing.	This step would increase the cost of testing (and is relevant only to shipboard installations and testing), so before it is required, the cost and benefit of a class society's involvement would need to be evaluated.

Table 25. Suggested changes to the Environmental Technology Verification Protocol (Continued).

Type of suggestion	TF(s)	Recommendation by TF	Comment by the VO
Sample collection and analysis	GBF GSI	Guidance on the diameter of sample ports should be consistent throughout the ETV Protocol, and the sizing of the ports should be revisited.	The ETV Protocol should be updated to ensure the guidance on sample ports is consistent throughout the document. The size of the ports (i.e., 1.5 – 2.0 x the isokinetic diameter) has generated much discussion and should be revisited.
	GBF	Prioritize the sample analyses described in the ETV Protocol so the highest priority samples are counted first and testing is less onerous; allow for changes to ETV Protocol using a 'Redline procedure' (changes during testing), as the ETV 'does not readily lend itself to being performed as a laboratory or strict production line process'.	The ETV Protocol should be re-evaluated and updated periodically, and any measurements shown by TFs' data to be superfluous should be removed. Any redlines to the SOPs should be approved by the VO. As far as assembly-line testing, the purpose of the ETV program in general is to ensure consistent approaches and standard methods are used at all facilities to generate reliable and comparable data.
	GSI	Provide guidance on counting treated water with numerous organisms, that is, how many organisms should be counted?	The guidance should be provided.
	GSI	Re-examine the use of in-line sensors, as they proved unreliable.	Each TF should validate its data-collection devices. TFs should use calibrated in-line sensors; doing so generally provides more reliable and comprehensive data than can be collected by hand.
	GSI	Require TFs to validate the holding time for samples of organisms (6 h is discussed in the ETV Protocol).	The ETV Protocol indicates the 'maximum hold time should be validated at each TF so that the detectable zooplankton mortality over the hold time does not exceed 5%' (Section 5.4.6.4 Zooplankton Enumeration). To ensure the results are valid, the 'should' ought to be 'shall'.
	GSI	Analyze organisms by taxonomic category rather than size limit, e.g., include rotifers (zooplankters) slightly less than 50 µm in the total of organisms in the size class of ≥ 50 µm.	Because the national and international discharge standards are based on size classes, data must be collected in this manner. An option is also to partition the data by taxonomy to provide insight into interactions between treatment and biology.
	GSI	Hold media blanks for organisms < 10 µm for 3-5 days prior to use to determine if contamination occurred.	The blanks should be held for > 1 day; the ETV Technical Panel should be consulted to provide guidance on this point.
	GSI	Provide more guidance on WET Testing and require such tests for BWMS using UV light (note that WET tests were not required in the Intercomparison Project).	This section of the ETV Protocol should be expanded or provide more references; regarding UV systems, since they are widely regarded to not produce active substances, ecotoxicity testing seems superfluous.
	GBF	Allow alternate approaches to be used to determine viability, e.g., for BWMSs using UV light.	The ETV Protocol is intended to be a living document, and as new approaches are validated, they should be included. Regardless, viability needs to be determined immediately prior to discharge according to the IMO and USCG discharge standards.

Table 25. Suggested changes to the Environmental Technology Verification Protocol (Continued).

Type of suggestion	TF(s)	Recommendation by TF	Comment by the VO
Guidance for the verification report	GSI	Provide additional guidance regarding the outline, appendices, information required in each section of the report, and DQI for engineering.	More guidance is necessary, and a sample report, template, or series of checklists should be available to TFs. They would provide a template for the structure of the report and its appendices, the inclusion of data in the validation matrix, a format for the data plots (to allow easy comparison among reports), and it would make more information available for the DQIs in general. Likewise, the guidance would define standard terminology; in that manner, TFs would use consistent terminology rather than using, for example, 'intake and discharge' or 'uptake and discharge' to describe the same ballasting operations.

BWMS = ballast water management system, DOC = dissolved organic carbon, DQI = data quality indicator, ETV = Environmental Technology Verification program, GSI = Great Ships Initiative, GBF = Golden Bear Facility, VO = Verification Organization, O&M = operations and maintenance, POC = particulate organic carbon, TF = Test Facility, UV = ultraviolet, VO = verification organization, and WET = whole effluent toxicity.

7.2 Additional Suggestions to Improve the Protocol

In addition to the suggestions from TFs for improving the ETV Protocol (Table 25), several additional ideas from the VO are discussed below. Items are discussed in an order that reflects the chronology of this project. First, aspects of testing are discussed, such as commissioning and the holding time of the uptake water. Next, the need for a template verification report is discussed.

7.2.1 Additional Suggestions to Improve the Protocol—Aspects of Testing

During commissioning of the BWMS, the ETV Protocol states it is 'not necessary to conduct a complete suite of analytical measurements to assess biological treatment efficacy.' If possible, however, it would be most useful to both the TF and the VO, particularly TFs new to testing a BWMS, to run a complete BE test as a practice run. Quantifying living organisms in treatment discharge samples during commissioning will also allow the TF to adjust sampling and analytical protocols, including determining the appropriate statistical approach. Given the demonstrated usefulness of performing pressure tests prior to commencement of testing (i.e., leaks were detected at both TFs during pressure testing), it is highly recommended that both a hydrostatic pressure test (at maximum recommended working pressure) and a hydrodynamic pressure test (at maximum expected operational pressure) be performed during the commissioning. Regardless, the commissioning report should be included as an appendix to the verification report.

The ETV Protocol does not address measuring the flow rate using a BWMS with a backflushing operation (i.e., any BWMS employing a filtration step). In future BWMS verification testing, it seems reasonable to use the approach from this project, that is, the flow rate should be calculated as the average flow rate excluding backflushing, unless vendor claims state otherwise. Likewise, stripping of control and treatment tanks is not addressed in the Protocol. It seems reasonable to allow TFs to slow deballast flows (following flow rates specified in their TQAP) at the end of drain operations to maintain suction and pump out as much water as possible.

The ETV Protocol also does not offer an upper bound on the duration of the water's hold time, only that it must be held for at least one day. Likewise, vendors generally do not specify a hold time for their BWMS, and they should do so. In this project, it was not feasible to conduct tests with a one-day hold time, as analyzing the uptake samples would have precluded making the necessary preparations to collect and analyze all of the control discharge and treatment discharge samples. Instead, a two-day hold time was used by both TFs, and the requirements for living organisms in the control discharge samples were met. When ETV Protocol testing is instituted, it would be difficult, for example, to compare results from one TF with a one-day hold time to results from a TF using a 10-day hold time. It seems unlikely such an extreme scenario would arise, but nonetheless, to avoid ambiguity, it is recommended that the ETV hold time be changed to at least one day and no more than five days (the latter is consistent with the IMO G8 guidelines). Finally, to avoid confusion, the ETV Protocol should define the 'hold time'; in these tests, the clock started as the tanks started to be filled and ended as each tank started to be drained. This approach seems reasonable.

Regarding communication between the VO and TF, during testing, the weeks were very busy, and holding a brief meeting (in person or by teleconference) after each test was useful. These conversations allowed both organizations to discuss any issues and anticipate any difficulties in the next test. When the role of the VO is described in the ETV Protocol, this idea should be added. Furthermore, the requirements, expectations,

and perceptions for the independence of TFs from the BWMS vendor and other interested parties are not explicitly stated in the ETV Protocol. These relationships should be explained in the ETV Protocol, or a reference to such information should be provided.

During testing, multiple analyses must be conducted within a relatively short period, which can strain a small testing team. Thus, it is imperative that the TF employ enough qualified members to safely and accurately collect and analyze samples. Ideally, the number of staff needed will be evident as SOPs are validated prior to ETV testing. Although the ETV Protocol addresses the testing personnel as "Key personnel and their organizations shall be identified, along with the designation of responsibilities" (Appendix A in EPA, 2010), a list of qualifications for each team member should be included in the Test Plan as well as a timeline for testing (as per GSI's "TQAP Appx 3 - Example Test Cycle Plan_Revision" within Appendix 1 [the TQAP] of the GSI verification report [Great Ships Initiative, 2012]). Additionally, QA/QC staff should be available to document deviations or redlines to SOPs so this burden is not added to personnel directly involved in test activities.

Further, it is tempting to conduct research on new methods while testing occurs. The ETV Protocol should discuss this scenario and make clear that the top priority must be collecting the "core" and "ancillary" measurements, with the new methods addressed only after all other work is completed. It is also imperative that SOPs are followed and DQIs are met to allow confidence in the data collected. Without such rigor, the results are questionable at best and unusable at worst.

During sampling, the condition of plankton nets or other filtration devices is critical. Given that capture of living organisms is central to ETV testing, a method for ensuring nets are in proper condition prior to each test should be included in the ETV Protocol (e.g., GSI's SOP 'Procedure for Cleaning Sampling Equipment at the GSI Land-Based RDTE Facility', [GSI/SOP/LB/G/O/4], found in Appendix 6 within Appendix 1 [the TQAP] of the GSI verification report [Great Ships Initiative, 2012]). Furthermore, any new equipment in contact with biological samples should be checked prior to use.

As GBF mentioned in their verification report, during sampling, some volume of the control and treatment tanks, approximately 1-3%, cannot be removed by the ballast pumps (and thus analyzed; Golden Bear Facility, 2012). It would be useful for the ETV Technical Panel to discuss ways to ameliorate this problem, but at a minimum, this volume should be quantified and mentioned in the verification report. Additionally, cleaning the ballast water piping and tanks is not explicitly discussed in the ETV protocol; both GBF and GSI have SOPs for cleaning tanks and piping which could serve as a starting point for developing a standard procedure(s) included in the ETV Protocol, and an inspection result should be specified in a TF's SOPs.

7.2.2 Additional Suggestions to Improve the Protocol—Sample Collection and Analysis

In Table 9 "Sample Volumes, Containers and Processing" of the Protocol, the minimum sample volume for the ≥ 10 and < 50 µm size class is 3 m^3, whereas later in the ETV Protocol (Table 14 "Sample Volume Required Relative to Treatment Standards—Organisms ≥ 10 and < 50 µm"), volumes of 2-6 L are discussed. This discrepancy should be addressed. The TFs participating in the Intercomparison Project obtained whole-water drip samples with final volumes of ~ 60 L (three 20 L carboys at GBF) and 15-47 L (up to three 19 L carboys at GSI), which ESAC and the TA agreed was sufficient to analyze organisms in this size class.

Although the Protocol discusses at length the statistical aspects of sampling the sparse populations in treated samples, the number of subsamples necessary from the uptake and control tank discharge is not addressed.

Ideally, the data collected by any TF is easily comparable to that collected from other TFs. To accomplish this ideal, it is suggested that a power analysis be conducted by each TF to determine the appropriate number of samples. In that absence of such analysis, it is recommended that at least five subsamples are collected for measurements of organisms in the two largest size classes. Furthermore, the Protocol does not provide guidance in the event that treated samples contain more living organisms than anticipated, as occurred in this project. It would be useful for the Protocol to outline a procedure for changing the analysis approach from one with sparse populations to one with non-sparse populations.

As the ETV Protocol currently reads, the analysis of organisms ≥ 50 µm is open to interpretation. That is, organisms should be examined for "at least 10 seconds for visible movement" (Section 5.4.6.4 "Zooplankton Enumeration", EPA, 2010). If some TFs examine organisms longer than others do, they could detect higher numbers of living organisms than other TFs. It is suggested that an upper bound of 20 seconds be placed on the examination window.

7.2.3 Additional Suggestions to Improve the Protocol—Data Collection and Management

A central tenet of the ETV program is the use of DQIs, as adherence to them provides confidence in the quality of data collection, sample analysis, and data reporting. The ETV Protocol provides brief statements on the DQI requirements for representativeness, accuracy, and precision, but outside of the glossary, it is essentially silent on DQI definitions of bias, comparability, and completeness. Despite limited guidance, both TFs did address many DQI requirements, data objectives were met in most cases, and those outputs can be used to improve verification reporting and TF SOPs under the ETV Protocol. Although the use of DQIs is ultimately specified through the TF's TQAP, improved guidance should be provided in the ETV Protocol to provide guidance for the use of multiple indicators and corresponding objectives, for biological, water quality, and engineering data. In this manner, QA/QC spikes, blanks, and duplicate analysis will be more widely applied when appropriate, and the requirements for operational DQIs will be better defined and applied.

Although it is not currently a requirement in the ETV Protocol, each TF should be required to develop a standardized naming system for all files with revision numbers; a system could be imposed by the ETV Protocol. The VO was able to complete a review of all documents, but at times was unsure if the most recent revision of a document was being reviewed. For future testing, the VO recommends that TFs develop and document data management practices by (1) documenting the steps for entering data into spreadsheets and checking the data entry, (2) developing data sheets when at all possible (rather than creating tables by hand in data notebooks), and (3) explaining the responsibilities of the QA/QC officer in reviewing the data. The TFs should develop a form to log all deviations so the QA/QC officer can track deviations, discuss them with the person in charge, determine how the deviation affected the results of the test, and provide appropriate corrective or preventive actions. The TFs should also describe the databases they will use and develop a systematic procedure to archive them (rather than writing a generic description). Regular communication among independent test facilities to exchange ideas would help all facilities to implement robust procedures.

7.2.4 Additional Suggestions to Improve the Protocol—Reporting

One confusing aspect of the ETV Protocol, which was evident during the preparation for testing, was the relationship between the QAPP and the TQAP, as well as the QMP and the Environmental Health and Safety Plan. It would be helpful if the ETV Protocol more explicitly explained each document and its use in

testing. While EPA websites are available, they may be of limited use in this context, given the complexity and unique nature of BWMS testing.

More guidance is needed regarding the format and content of the verification report (as in Table 24 and Table 25 above). Because of the lack of clarity in the requirements for reporting in the current ETV Protocol—as well as the newness of ETV testing to both the TFs and the VO—each TF provided three iterations of the verification report to the VO and AO for their review—the TFs' persistence is commendable! Additionally, the ETV Protocol should require tables of 'deviations' from the SOPs and the effect on test outcomes are included in the verification reports.

Over the course of this project, it was apparent that some inconsistencies in the ETV Protocol should be addressed in its next publication, which would assist in the preparation of verification reports and allow easy comparison among different reports. One was the use of units: parts of the document used the International System of units for measurements, typically for laboratory measurements of volumes, and other parts used the U.S. system of units for measurements, typically for engineering measurements, such as pipe diameters. The VO recommends listing both units in reports when it is possible that users of the ETV Protocol (or other stakeholders interested in testing outcomes) could subscribe to either system.

7.3 Comparison of the Environmental Technology Verification Protocol and the International Maritime Organization G8 Guidelines

Because the ETV Protocol was finalized after the IMO G8 guidelines for sampling, the ETV Protocol benefited from insights gained by the scientific and regulatory communities in the interval between the publication of the G8 guidelines and the publication of the ETV Protocol. Primary differences between the documents are listed in Table 26. Importantly, the sample volumes for organisms ≥ 50 µm are much larger in the ETV Protocol than in the G8 guidelines. The two documents differ in other ways: the challenge water conditions are dissimilar, the number of BE tests per salinity is lower in the ETV Protocol (3) than in the G8 guidelines (5) (although the USCG Final Rule specifies 5 tests per salinity [USCG, 2012]), and the hold time specified in the ETV Protocol (≥ 1 day) is less than that in the G8 (≥ 5 days). Regarding the hold time, this issue would (and should) be easily addressed if vendors specified the hold time required to meet the BWMS' claims of efficacy. Overall, the ETV Protocol contains more detailed instruction, particularly concerning methods for biological analyses, requirements for data quality assurance, and contents for the test plan and verification reports than the G8 guidelines. The more detailed guidance provided by the ETV Protocol will result in more comparable data sets and test reporting across facilities than that produced by the various interpretations of the current G8 guidelines, which is silent on many of these details.

Table 26. Requirements in the Environmental Technology Verification Protocol and International Maritime Organization G8 Guidelines.

Organization and Guidance	Metric						
	Living Organisms ≥ 50 µm in minimum dimension	**Living Organisms ≥ 10 µm and < 50 µm in minimum dimension**	**Toxigenic *Vibrio cholerae***	**Coliform and enterococci group**	**Living Organisms < 10 µm in minimum dimension**	**Use of ambient organisms**	**Number of tests DOC, POC, TSS Temperature Salinity**
US ETV Protocol "Challenge water" for land-based testing[A] 3 valid tests at 2 salinities ≥ 1-day hold time con and trt tanks ≥ 200 m^3 and flow rate ≥ 200 m^3 h^{-1}	≥ 100,000 m^{-3} ≥ 5 species from ≥ 3 different phyla Control tank discharge: ≥ 100 m^{-3} Sample volume: ≥ 3 m^3; also says it is the TF's choice	≥ 1,000 mL^{-1} ≥5 species from ≥ 3 different phyla Control tank discharge: ≥ 100 mL^{-1} Sample volume : ≥ 3 m^3; also says it is the TF's choice	Not addressed other than method described to measure concentration	Not addressed other than method described to measure concentration	1,000 mL^{-1} culturable, aerobic heterotrophs Discharge: Control tank ≥ 500 mL^{-1}	STOs only required in preliminary bench-scale testing	DOM: 6 mg L^{-1} (as DOC) POM: 4 mg L^{-1} (as POC) MM: 20 mg L^{-1} TSS: 24 mg L^{-} T = 4-35 ºC <1, 10-20, 28-36 psu
IMO G8 "Influent water" for land-based testing[B] 5 valid tests at 2 salinities ≥ 5-day hold time con and trt tanks ≥ 200 m^3 and flow rate ≥ 200 m^3 h^{-1}	Preferably 1,000,000 m^{-3} but not less than 100,000 m^{-3} ≥ 5 species from ≥ 3 different phyla or divisions Control tank discharge: ≥ 100 m^{-3} Sample volume: ≥ 20 L of influent ≥ 1 m^3 of treated	Preferably 10,000 mL^{-1} but not less than 1,000 mL^{-1} ≥ 5 species from ≥ 3 different phyla or divisions Control tank discharge: ≥ 100 mL^{-1} Sample volume: ≥ 1 L of influent ≥ 10 L of treated	These organisms do not need to be added at the influent but should be measured at influent and discharge: Coliform bacteria Enterococcus group *Vibrio cholerae* Heterotrophic bacteria		At least 10,000 mL^{-1}; note that this guidance is incongruent with the cell to the left	Test organisms maybe naturally occurring or cultured species	At least 2 test cycles (each cycle = 5 reps) completed at 2 different salinities, separated by 10 psu if within adjacent salinity types > 32 psu/0-32 psu DOC: > 1/> 5 mg L^{-1} POC: > 1/> 5 mg L^{-1} TSS: > 1/> 50 mg L^{-1} T range is unspecified < 3, 3-32, > 32 psu

[A] Criteria are to be met (1) at influent point of control tank (2) point of treatment or treatment tank influent point. [B]Samples are to be collected in triplicate immediately before treatment equipment, immediately after treatment equipment, and upon discharge from treatment tank; control samples are to be collected upon influent and discharge. cfu = colony-forming unit, con = control, DOC = dissolved organic carbon, DOM = dissolved organic matter, ETV = Environmental Technology Verification program, FW = freshwater, MM = mineral matter, POC = particulate organic carbon, POM = particulate organic matter, psu = practical salinity unit, reps = replicates, STO = standard test organism, T = temperature, TF = test facility, trt = treatment, and TSS = total suspended solids (= POM + MM).

Regarding the differences between the two approaches for testing BWMS, in most cases, it is premature to say which approach is "better". For example, although the requirements for dissolved and particulate matter differ, until efficacy data are available for type approved BWMS (How well do the systems work after they have been type approved?), one cannot say if the requirements are too stringent or not stringent enough. Likewise, whether three tests at each salinity (as per the ETV Protocol) or five tests (as per the G8 Guidelines) are sufficient to predict the real-world performance of a BWMS will be (hopefully) clear after performance data are collected from the global fleet of vessels using treatment technologies. Importantly, the ETV Protocol allows for changes to requirements with the approval of the VO. Likewise, the G8 Guidelines can be modified as deemed necessary by a party to the convention.

One difference, however, is worth mentioning: the sample volumes for organisms ≥ 50 µm in the treatment discharge samples. The scientific community has largely come to consensus that relatively large volumes (m^3 to tens of m^3) must be collected in a manner that is representative of the entire volume of interest to determine with statistical confidence the number of living organisms (e.g., Lee et al., 2010; Miller et al., 2011; U.S. EPA, 2012). Thus, the ETV Protocol's guidance for sampling this largest size class improves measurement accuracy over the G8 guidelines, which were drafted prior to the recent assessments of appropriate sample sizes and statistical approaches for evaluating the efficacy of BWMS.

Irrespective of differences between the two approaches, the outcome of this project's BE testing shows that the BWMS tested discharged organisms in excess of the IMO standard at both test facilities. A comparison between the data and information collected during Type Approval testing with those of this effort is beyond the scope of this project. Therefore, it is impossible to extend a more rigorous comparison between test outcomes, but at a minimum, the hold time differed between tests (five days were used in the Type Approval tests, following the IMO G8 guidelines; two days were used in this project), which likely affected the concentration of living organisms in treatment discharge water.

8 CONCLUSIONS

The goals of this effort were to objectively evaluate the comparability of BWMS verification testing and quantify the variability between test facilities. To this end, the TFs' capabilities to apply the methods and procedures of the ETV Protocol, and where possible, the IMO G8 Type Approval Guidelines were assessed. The objectives to accomplish these goals were to choose and purchase a commercial-off-the shelf BWMS; use a recently developed questionnaire to assess TFs' qualifications and solicit TF participation; devise a rubric for scoring TFs' responses; choose two TFs to participate in the project; advise the facilities on the preparation of all necessary documentation (test plans, etc.); and conduct three, valid BE tests at each of the two TFs, yielding two verification reports. All objectives were successfully met. That said, some questions were raised by this effort.

Can TFs conduct verification testing using the ETV Protocol?

The answer to this important question is not straightforward. This work showed yes, testing can be done in a reliable, repeatable manner using the methods, procedures, and QA/QC measures outlined in the ETV Protocol. Notably, because this was a research project and the timeline was tight for GBF to prepare their facility for testing, a number of modifications or substitutions to the ETV Protocol were approved without prior validation. It is anticipated that such modifications would have the appropriate documentation and validation for true ETV testing. Both TFs operated their facilities with adequate control—either largely manual (GBF) or largely automated (GSI)—to obtain the target flow rates, tank volumes, and sample

volumes specified in the ETV Protocol. Importantly, both TFs were able to document and verify test conditions over the duration of each test. Both TFs executed a TQAP, developed a QMP, and presented final test data in a comprehensive verification report.

This work also showed no, testing cannot be done if all challenge water conditions in the ETV Protocol must be met. Even with testing judged according to the DFS granted to GBF of lowering the concentration of living organisms ≥ 50 µm, one of three tests met the uptake requirements for all three size classes. Likewise, even with the DFSs granted to GBF for water quality parameters, none of the tests met all water quality conditions (DOC, POS, and TSS), and GSI met the conditions in only two of three tests. In judging the tests with no DFSs, that is, using the ETV Protocol requirements for water quality and organisms, neither TF met all requirements in *any* of the tests. Regarding water quality parameters, GBF, which did not have the capacity to augment ambient waters, could consistently reach approximately one-third of the required concentrations of POC and DOC, although MM concentrations were exceed by more than two- to three-fold. GSI, which did augment ambient waters, nearly always met the water quality requirements, and in the case of DOC, doubled or tripled the required concentration.

Achieving challenge water requirements for living organisms proved difficult at the TFs. Although GSI augmented ambient organisms in the ≥ 10 µm to < 50 µm size class, the requirement (measured according to the ETV Protocol, not the DFS) was not met in any BE test, reaching approximately 60% of the required concentration. The requirements for the other two size classes, meanwhile, were exceeded at GSI by approximately 3- to 4-fold (≥ 50 µm) and 10- to 190-fold (< 10 µm). GBF did not meet the challenge water conditions for the ≥ 50 µm size class in any test (measured according to the ETV Protocol, not the DFS; 20-90% of the target concentration was attained) or two of three tests for the ≥ 10 µm to < 50 µm size class (reaching approximately 30-80%) but nearly doubled the requirement in one of the tests. For organisms < 10 µm, the requirement was exceeded at GBF approximately two-fold in the first two tests and very nearly met in the third. These outcomes show either the TFs will need to augment the challenge water (GBF) or change their means to augment challenge water (GSI), or the ETV challenge water requirements need to be revisited. Of course, any augmentation procedure should be appropriately validated, and one consideration should be the potential for biological communities to change seasonally at TFs, which may require multiple validation exercises.

If these two TFs had difficulty meeting the challenge requirements, other facilities, both in the U.S. and abroad, likely face the same problems. This issue should be forwarded to the ETV Technical Panel to consider lowering some requirements or recommending guidelines for acceptable "misses". Regarding the latter, for example, it might be permissible for 75% of each requirement to be met in 100% of the BE tests (e.g., in lieu of 100,000 organisms ≥ 50 µm m^{-3}, 75,000 m^{-3} would be acceptable). Conversely, it might be permissible to meet 100% of the requirements in only 66% (2 of 3) of the BE tests. Regardless, for regulators and other stakeholders to assess the testing data, a re-examination of the requirements for challenge water conditions or a re-examination of the means to augment challenge water is merited.

In this project, two types of variability can be assessed: within facility and between facility. Examining the mean uptake concentration of living organisms in the two largest size classes *within* tests at a TF, the concentrations varied at most by a factor of six. That variability is neither surprising nor high for biological data collected in the field. At each TF, all thee BE tests were completed in three or four weeks, and it is likely that the ambient community changed during that time (or over a matter of days or hours, for that matter). Examining the data from the control discharge and treatment discharge for the two largest size classes, data varied within a factor (control or treatment discharge) at most by three fold across BE tests

within a facility. The counts of organisms < 10 µm at GSI and GBF differed within a factor (control or treatment discharge) by a factor of 15 and 38, respectively (the latter was potentially due to the items discussed by NSF). The results from the smallest size class notwithstanding, these outcomes are reassuring, both from the facility's standpoint—as they illustrate the TFs can individually run repeatable tests—and from the ETV Protocol's standpoint—as they illustrate the Protocol's use as a guide for TFs.

Despite differences *between* TFs, are the outcomes comparable?

Ideally, the answer is yes. In this project, the concentrations of living organisms in the three size classes collected in uptake, control discharge, or treatment discharge samples were often significantly different among tests between the facilities. In no case, however, were the mean values of all three tests at one TF significantly different from the mean values of all three tests at the other TF. Additionally, although the concentrations of organisms ≥ 50 µm in treatment discharge samples from both TFs differed significantly among tests, and the number of organisms was reduced (and with a similar magnitude) between uptake and treatment discharge water, and the number of living organisms in each trial in treatment discharge exceeded current IMO and USCG discharge standards. Again, this result points to the repeatability of the Protocol, even between locations with different biological communities and water quality characteristics. The treatment tank data from organisms ≥ 10 µm and < 50 µm, however, yielded different results between facilities. Despite the initial concentrations being rather similar (concentrations ranged from approximately 300-2,000 organisms mL^{-1} at GBF and 550-600 organisms mL^{-1} at GSI), the concentrations from all three tests at GBF were low and as much as 45 times lower than the aggregate mean of all three tests at GSI. It seems the differences in community composition between locations are needed to effectively assess treatment efficacy for a given size class, although it is possible the largest organisms (≥ 50 µm) clogged the filters at GSI, reducing the efficacy of the UV treatment on the size class for organisms ≥ 10 µm and < 50 µm. From this study, it is not possible to untangle these potential interactions. Nonetheless, this result highlights the importance of determining the efficacy of a BWMS using different size classes of organisms. It also illustrates the importance of conducting land-based validation testing at multiple salinities across the range of salinity regimes where the system is intended to be used, which will contain different communities. Furthermore, shipboard validation testing will challenge BWMSs with different communities, even if the challenge conditions are less controlled, lower, or both, than those used in land-based testing.

How do differences between facilities' physical arrangements that warrant different practices affect the test results?

An example of such differences is the stripping of the ship's ballast tanks used for testing at GBF vs. suspending organisms at GSI with a mixer. Another is using one sample port at GBF vs. using three sample ports at GSI. These differences should not matter if the practice does not induce mortality upon organisms and if the samples collected are representative of the volume of interest (here, the entire tank). Regarding mortality, the largest size class would be more likely affected by turbulence-inducing practices than the smallest size classes, which are influenced more by viscous forces than inertial forces (e.g., Fenchel, 1987). In this project, the concentration of large (≥ 50 µm) organisms in the treatment discharge samples were relatively high, so no effect was evident. Still, it is incumbent on each TF to demonstrate its practices do not bias the data. Regarding the representativeness of the samples, both TFs used sample ports following the guidance in the ETV Protocol, so, despite the error that is inherent in sample collection and handling, there is reason to assume the samples were representative of the volume discharged. However, regarding residual volumes in tanks, if a disproportionately high or low concentration of organisms is found in the water remaining in the tank as compared to the water sampled upon discharging a tank, the

representativeness of the sample would be called into question. Thus, to validate the representativeness of samples, the volume of residual water should be quantified and the concentration of organisms determined. Regardless, TFs should strive to minimize residual volumes of ballasted water.

A comparison of the major sources of error between facilities is difficult to make when the biological community concentration and composition of the uptake challenge water conditions differ between TFs. A robust way to conduct a comparison would be to prepare uptake water for each facility with the same STOs at the same concentrations. This approach, however, is unappealing. First, the STOs, from freshwater and from saltwater, for each of the three size classes, would need to have similar tolerances to treatment within a size class. Finding such organisms would be very difficult. Next, huge volumes (200 m^3 or 400 m^3) of ambient water would need filtering to remove ambient organisms, or huge volumes of artificial seawater would need to be prepared. The STOs would need to be cultured in quantities great enough to be added to the water to meet challenge water conditions. The last (and potentially largest) obstacle would be to process the water after testing to certify that non-native STOs were not inadvertently released to the environment. This thought experiment shows this approach to determining error between TFs to be untenable. Alternatively, blanks or spikes (e.g., a local organism) might be routinely incorporated into test protocols, allowing for intra- and inter-facility comparisons and serving as additional data quality metrics. Finally, taking a theoretical approach, it is possible to create a mathematical model of error across a facility by considering all steps in which measurement error can be made (e.g., measuring the volume of concentrated sample or the volume of sample analyzed), thus using empirical data to develop and test the model; that work is underway.

Do differences among TFs create an opportunity for a BWMS vendor to obtain an unfair advantage by choosing a test site that favors his or her system?

Certainly, the ability to meet minimum challenge conditions is requisite to minimizing differences among facilities, and it may require TFs to augment their ambient water for testing under the ETV Protocol. Discrepancies in source water conditions across facilities will be unavoidable, and, to the extent that they reflect variability in the real world, can be desirable. Given their baseline of source water conditions, the traditionally high dissolved organic load at GSI may pose a more significant challenge for some technologies such as those using UV light. Whether this situation is common with other freshwater facilities (or saltwater TFs, for that matter) remains unknown, but such factors may influence vendors choosing a site for verification testing. One way to address this issue is to set threshold challenge conditions based on the type of treatment employed by a BWMS. For example, UV systems would be tested in water with low transmissivity, and oxidant systems would be tested in water with high DOC and POC concentrations. At present, such a matrix of treatment types and conditions does not exist. The logical body to develop such a framework would be the ETV Technical Panel.

Another way to address this issue is by including in the verification report a context for the challenge conditions in the tests. That is, a discussion would be included regarding the representativeness of the conditions—are they encountered in harbors and coastal waters, and do they present a robust challenge for the BWMS based on the technology's capabilities or the results of laboratory testing? If not, by what magnitude do the conditions exceed or fall short of those benchmarks? These suggestions notwithstanding, in this project, the similarity in the overall test results from the two facilities suggests that high DOC levels at GSI were not the only challenge the BWMS encountered.

A goal of this project was to evaluate the ability of U.S. TFs to conduct validation testing in a consistent and comparable manner. Considering the overall complexity of the ETV Protocol and its recent introduction, one would expect there to be a learning curve for some of the many requirements. In particular, the strict documentation, auditing, and quality assurance procedures required for standardized testing may present a challenge to traditional research teams now making the change from experimental testing to standardized testing.

In this project, the imposition of challenge conditions, sampling volumes, analysis protocols, and validation required the TFs to change existing protocols or develop new ones. Participation required VO and AO review and concurrence of many documents prepared by the TFs: documents regarding quality management, environment health and safety, staffing and training records, as well as a comprehensive TQAP. None is trivial to prepare, illustrating the commitment of each facility to testing. At GBF, a new facility now performing land-based testing on a ship, many protocols were modified or written specifically in response to the requirements of this project. There were new instruments, equipment, and procedures that had not been previously exercised or used extensively under test conditions. Most staff conducted multiple jobs, and the biology staff were on-site, analyzing samples, until late in the night. This situation, if part of a TF's normal procedure, demonstrates more staff are needed, and in fact, GBF has stated an intention to add more staff. At GSI, which had a longer history of conducting similar tests, and which has increased in staff size over the recent past, staff members were assigned to single jobs, and their tasks were finely choreographed (the GSI Test Cycle Plan included tasks as brief as five minutes long). As a consequence, although the staff worked long days, they did not need to work late into the night—such an arrangement should be the goal of all TFs.

The biological methods referenced in the ETV Protocol were developed over years with the final goal of being used in multiple locations. The methods worked at both of the TFs, one in freshwater, one in brackish water, in that both facilities employed the methods outlined in the protocol, generated consistent results among tests at each facility, and successfully met most data quality metrics. The methods also yielded data expressed as concentrations of organisms per unit volume, rather than, for example, the mass of organisms per unit volume. This outcome is important and relevant because the U.S. and IMO discharge standards are expressed as densities of organisms. The results of this project are promising for this type of evaluation.

One outcome not assessed to the desired extent was the measurement of rare populations of organisms ≥ 50 µm in water treated by a BWMS. Both TFs demonstrated appropriate sampling methods (i.e., collecting and processing relatively large volumes of water) for rare populations. However, the concentrations of organisms measured at both TFs were well above the densities for which the statistical guidance for treated samples in the ETV Protocol is appropriate. In consequence, all discharge samples for organisms ≥ 50 µm at both TFs were analyzed as control samples (i.e., subsampled because they had a high density of organisms with an assumed statistically normal distribution). This situation prevented a comparison of outcomes within or across facilities regarding the ability to resolve the concentration of organisms ≥ 50 µm in treatment discharge water at or below regulatory limits.

The application of DQIs is a requirement of the ETV Protocol—and DQIs are needed to ensure confidence in the data—but the ETV Protocol guidance is minimal, and the specific application is left to the TF to explain in its TQAP. For example, both facilities in this project applied DQIs to biological measures and to some extent water chemistry. Regarding engineering DQIs, GBF applied them to some engineering measures, while GSI did not. It is logical to extrapolate from this case study and assume other TFs would also interpret the ETV guidance in their own manner. Thus, more guidance for the appropriate use and

reporting of DQIs is required in the ETV Protocol. While the ETV Protocol indicates the facility should assess variability, examine bias, and incorporate spikes and blanks where appropriate, a draft verification report, template, or series of checklists regarding biological counting methods would serve facilities well. Engineering indicators should examine calibration data, compare readings at multiple locations, and assess variability in reported data. Completeness, while mentioned in the Protocol, is not applicable to logged engineering data, and representativeness is more qualitative than quantitative. Again, additional guidance would be helpful.

Another useful addition to the protocol would be a table listing the methods recommended in the ETV Protocol for biological and water quality analyses; the table would also include a blank column for TFs to populate with the methods used at the facility. Such a uniform presentation would allow easy comparison of the TF's methods with the ETV Protocol, and it would facilitate comparison among multiple verification reports. Furthermore, it is recommended that the community of TFs regularly meet to review procedures for evaluating the quality of the biological, water quality, and engineering data to ensure the data are unbiased, representative, and reliable. Such ongoing assessment and feedback would result in continuous improvement at all facilities. Finally, during testing, it is strongly recommended that all 'core' parameters are addressed first (e.g., completing the necessary sample handling and sample analysis) before beginning ancillary measurements or conducting additional analyses outside those recommended in the ETV Protocol. Although these recommendations were made based on observations of the two participants in the Intercomparison Project, they are applicable to any TF conducting testing under the ETV Protocol.

Finally, although a goal of this project was to understand and quantify the differences between two facilities, invariably, questions regarding the efficacy of the BWMS compared to U.S. or IMO standards will arise. Under this Intercomparison Project, the BWMS under test was considered a generic piece of ballast water management equipment that could be installed at two domestic TFs and used to compare their execution of the ETV Protocol. Because not all aspects of the ETV Protocol were tested, such as O&M testing, and because the BWMS vendor was not fully involved in the development of the TQAP and did not review the verification report, this research project was not an ETV test. Thus, the results should not be considered ETV data. More broadly, this project could not compare the results from this testing to results reported for this or other BWMSs tested elsewhere, as such a comparison was outside the scope of the project. Further, such a comparison to other tests would be extremely difficult and of questionable value without the detailed descriptions of methods and QA/QC procedures used. Indeed, it is likely that the current lack of consistency in test methods among TFs testing BWMSs globally makes such a comparison impossible at this time. Such consistency in test methods and test quality is the fundamental purpose of the ETV Protocol.

9 REFERENCES

Fenchel, T.D. (1987) *Ecology of Protozoa*. Science Tech, Inc. Madison.

Golden Bear Facility (2012) Demonstration of Environmental Technology Verification Protocol for Ballast Water Management System Testing, California Maritime Academy, Vallejo CA.

Great Ships Initiative (2012) Great Ships Initiative Verification Report for Land-Base Test of the Government Furnished Ballast Water Treatment System, Great Ships Initiative, Superior, WI.

International Maritime Organization (2004) Convention BWM/CONF/36 'International Convention for the Control and Management of Ships' Ballast Water and Sediments, 2004'.

International Maritime Organization (2005) Resolution MEPC.125 (53) 'Guidelines for Approval of Ballast Water Management Systems (G8)'.

Johnson, R. A. and Bhattacharyya, G. K. (2006) *Statistics Principles and Methods* (5th ed.). John Wiley & Sons, Inc. Hoboken.

Lee II, H., Reusser, D. A., Frazier, M., and Ruiz, G. M. (2010) Density Matters: Review of Approaches to Setting Organism-Based Ballast Water Discharge Standards. *Report number EPA/600/R-10/031, United States Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division,* Corvalis, OR.

Lemieux, E.J., Grant, J.F., and Wier, T.P. (2010a) Ballast water test facility specifications for automated control of operations. Letter Report 6130/1026. Naval Research Laboratory, Washington, D.C.

Lemieux, E.J., Grant, J.F., Wier, T.P. and Drake, L.A. (2010b) Ballast water test facility questionnaire. Letter report 6130/1027. Naval Research Laboratory, Washington, D.C.

Lemieux, E. J., Robbins, S., Burns, K., Ratcliff, S. and Herring, P. (2008) Evaluation of Representative Sampling for Rare Populations Using Microbeads. *Report number CG-D-03-09, United States Coast Guard Research and Development Center,* Groton, CT.

Lenth, R. (2001) Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician,* **55**, 187--193

Miller, A.W., Frazier M., Smith G.E., Perry E.S., Ruiz G.M. and M.N. Tamburri (2011) Enumerating Sparse Organisms in Ships' Ballast Water: Why Counting to 10 is Difficult? *Env. Sci. and Technol.*, **45**, 3539--3546.

Reavie, E.D., Cangelosi, A.A. and Allinger, L.E. (2010) Assessing ballast water treatments: evaluation of viability methods for ambient freshwater microplankton assemblages. *J. Great Lakes Res*., **36,** 540--547.

Richard, R. V., Grant, J. F. and Lemieux, E. J. (2008) Analysis of Ballast Water Sampling Port Designs Using Computational Fluid Dynamics. *Report number CG-D-01-08, United States Coast Guard Research and Development Center,* Groton, CT.

U.S. Environmental Protection Agency (2011) *Efficacy of Ballast Water Treatment Systems: a Report by the EPA Science Advisory Board, EPA-SAB-11-009, United States Environmental Protection Agency*, Washington, D.C.

U.S. Environmental Protection Agency (2010) Generic Protocol for the Verification of Ballast Water Treatment Technology, Version 5.1. *Report number EPA/600/R-10/146, United States Environmental Protection Agency Environmental Technology Verification Program*, Washington, D.C.

U.S. Coast Guard (2012) Standards for Living Organisms in Ships' Ballast Water Discharged in U.S. Waters (Final Rule); *Federal Register*, **77**, 17254--17304.

U.S. Coast Guard (2009) Standards for Living Organisms in Ships' Ballast Water Discharged in U.S. Waters (Draft Programmatic Environmental Impact Statement, Proposed Rule and Notice); *Federal Register*, 74 44631—44672.

U.S. Coast Guard Research and Development Center (2009) *FBO Solicitation Number: HSCG32-09-I-R00015, Sources Sought: Ballast Water Treatment Test Facilities,* Groton, CT.

Whitlock, M. and Schluter, D. (2009) *The Analysis of Biological Data*. Roberts & Company. Greenwood Village.

10 APPENDICES

Three appendices (A-C) to this report are provided as separate electronic files to reduce the size and complexity of the report. A single page for each appendix is provided below; each provides a top-level list of contents for the corresponding appendix.

APPENDIX A. CONTENTS OF THE VERIFICATION REPORT FROM THE GOLDEN BEAR FACILITY, VALLEJO, CA

Title: Demonstration of Environmental Technology Verification Protocol for Ballast Water Management System Testing, Revision B

California Maritime Academy Department ID 76109
Dated 30 April 2012

Authors:
Nick Welschmeyer, PhD, Lead Scientist, Golden Bear Facility / Moss Landing Marine Laboratories
William T. Davidson, Facility Directory, Golden Bear Facility
Kevin J. Reynolds, PE, Marine Engineer / Naval Architect, Golden Bear Facility / The Glosten Associates

Table of Contents
Revision History
List of Tables
List of Figures
List of Abbreviations
References
Executive Summary with Validation Matrix
Chapter 1: Project Introduction and Background
Chapter 2: Ballast Water Management System Description
Chapter 3: Experimental Design
Chapter 4: Challenge Water Conditions
Chapter 5: Methods and Procedures
Chapter 6: Scientific Results – Biological Efficacy
Chapter 7: Quality Assurance and Quality Control
Chapter 8: Data Quality Indicators
Chapter 9: GBF ETV Conclusions
Chapter 10: Acknowledgements
Chapter 11: Appendices A-H
List of Appendices
Appendix A: Project TQAP with Redlines (Redacted) (I: Project Plan, II: Quality Assurance Procedures Plan, III: Standard Operating Procedures, IV: Vendor Operations Manual [only in non-redacted TQAP])
Appendix B: Data Logs – Engineering Standard Operating Procedures (181 pages of scanned forms and logs)
Appendix C: Data Logs – Automations System Outputs (551 pages of data plots, truth tables and raw data)
Appendix D: Data Logs – Water Quality Analysis Reports (McCampbell Analytical, 30 pages of laboratory logs)
Appendix E: Data Logs – Biology Notebooks (125 pages of scanned notes)
Appendix F: QA/QC Data and Figures – Biology Efficacy Raw Data Listings (61 pages of tables and charts)
Appendix G: QA/QC Data and Figures – Engineering Parameters (65 pages of data plots)
Appendix H: QA/QC Data and Figures – Water Quality Parameters (18 pages of data plots)

(This page intentionally left blank.)

APPENDIX B. CONTENTS OF THE VERIFICATION REPORT FROM GREAT SHIPS INITIATIVE

Title: Great Ships Initiative Verification Report for Land-Based Tests of the Government Furnished Ballast Water Treatment System, version #2

Great Ships Initiative, Northeast-Midwest Institute
Dated 04 MAY 2012

Principle Investigator: Allegra Cangelosi, NEMWI

(This page intentionally left blank.)

APPENDIX C. CONTENTS OF INTERCOMPARISON STATISTICAL OUTPUT

Title: Intercomparison Statistical Output

U.S. Naval Research Laboratory
Dated 30 MAY 2012

Authors:
Lisa A. Drake, US Naval Research Laboratory, Code 6136
Edward J. Lemieux, U.S. Naval Research Laboratory, Code 6130
Jonathan F. Grant, Battenkill Technologies, Inc.
Evan W.J. Parson, Vision Point Systems
Timothy P. Wier, Excet, Inc.

Overview:
The statistical output is broken down into three distinct sections: the Kruskal-Wallis pairwise comparisons performed with a Familywise Error Rate (FWER) of 0.20, the pairwise comparisons performed with a FWER of .05, and the Mann-Whitney U tests performed to compare the aggregate test data (GBF vs. GSI). Additionally, while the multiple comparisons charts are included in-line with the appropriate test and output, they can also be found in the Appendix in order to make quick visual comparisons between the tests performed with differing error rates.

Contents:

(This page intentionally left blank.)